Department of Physics and Astronomy
University College London
University of London

First-principles Studies of Surface Defects of Model Metal-Oxide Semiconductors

Nikolaos Beglitis

University College London

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of University College London

I, Nikolaos Beglitis, declare that the work presented in this thesis is my own. Where information has been obtained from other sources, I declare this has been clearly indicated in the thesis.

Abstract

In this thesis, three different model metal-oxide semiconductor systems will be discussed. First, the impact of hydroxyl vacancies, OH_{vac}, on the geometry, electronic structure, and mechanical properties of single-walled aluminosilicate, $(Al_2SiO_7H_4)_N$, and aluminogermanate, $(Al_2GeO_7H_4)_{36}$, nanotubes is investigated. It is found that, with the exception of one OH_{vac} localised on the outer wall of the $(Al_2GeO_7H_4)_{36}$ tube, these defects induce occupied and empty states in the band gap. Those states are found to be highly localised both in energy and in real space. Different magnetisation states are also found, depending on both the chemical composition and the specific side with respect to the tube cavity.

The focus of the thesis then shifts to one of the most important and well-studied metal-oxide surfaces, the rutile $TiO_2(110)$ surface. The reactivity of the surface is revisited, in view of the discrepancy between theory and experiment on the interaction between molecular oxygen and surface hydroxyls. This discrepancy is resolved by proposing that excess charge, associated with the oxygen vacancy and originating from Ti interstitials, is present on the surface. This surface charge opens new reaction channels not theoretically possible otherwise. The study utilises hybrid Density Functional Theory (DFT) calculations and Scanning Tunneling Microscopy (STM) simulations to provide evidence for the proposed surface charging.

The last part of the thesis focuses on another surface of TiO_2, the (011) surface. $TiO_2(011)$ has recently attracted attention due owing to its reported high photocatalytic activity. Several proposed structures of the surface are inconsistent with each other. Recent developments, based on Surface X-Ray Diffraction (SXRD) data and DFT simulations, now agree on a new structure. In this part a review of the various structures is provided and further evidence is given on the validity of the new proposal by providing further insight on the appearance of the surface on the STM.

Acknowledgments

I am particularly grateful to Prof. Andrew Fisher for his guidance and support and for giving me the opportunity to begin scratching the surface of such an exciting and intellectually stimulating field like materials modeling. His patience, encouragement and advice has been invaluable. In addition I would like to thank Dr. Gilberto Teobaldi for assisting me with the subtleties of DFT and STM simulations and for spending countless hours explaining the intricacies of VASP, bSKAN and materials modeling to me. I am also grateful to Prof. Geoff Thornton and his group for providing guidance regarding their experimental results and to Dr. Anthoula Papageorgiou for the endless discussions over coffee on TiO_2 past and current research activity.

In addition, I would like to thank the staff at the Physics Department of the University of Crete for their efforts to structure an environment where scientific knowledge is a worthy pursuit. By not making any compromises, they certainly made physics more appealing to me. More particularly, Prof. Theodore Tomaras has been a source of inspiration over time, his guidance in the late days of my undergraduate degree and the early days of my postgraduate study was invaluable and it is gratefully acknowledged.

I would also like to thank my friends, Giorgos Fragkos, Avgoustinos Filippoupolitis, Aggelos Karkalemis and Vasilis Korolis for all the fun years and Eleni Kostopoulou for, among many other things, her encouragement and support the last days of writing this thesis. My manager at Barclays Capital, Mathew Pitchforth, should be credited with allowing me, contrary to investment banking habits, to come late to the office after a night of hard work on the final drafts. Last but not least, I would like to thank my family for their endless efforts and support over the years, without any doubt things wouldn't be the same without them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Prologue

During the second half of the 20^{th} century, progress in theoretical physics and chemistry, coupled with advanced computational techniques and the advent of increasingly powerful digital computers, has given birth to the field of computational physics. It is now possible to perform highly accurate computer simulations which, alongside sophisticated experiments, can provide new insights in the behaviour of physical systems, ranging from low-energy lattice QCD to galaxy formation, and from the chemistry of single atoms or molecules, to ab-initio or first- principles studies of large bio-molecules, polymers and surfaces.

In materials science and solid state physics, metal-oxides have traditionally attracted significant attention. The variety of phenomena and the wide range of industrial applications, have made them one of the classes of materials most intensively studied using computational and numerical methods. Moreover, it has been established that many of their interesting properties are due to a departure from periodicity and the introduction of defects. Colour, catalysis, corrosion, and many more are just some of the phenomena that are facilitated by the presence of defects, either in their bulk structure or the surface. It is therefore entirely anticipated that the already established momentum in active research on the electronic properties of various metal-oxides and their surfaces, will continue to grow.

In this thesis, some of the theoretical and computational techniques used in quantum mechanical modeling of materials will be discussed, along with their application for the

study of two model metal-oxide semiconductors. More particularly, the electronic structure of pristine and defected aluminosilicate (AlSiOH) and aluminogermanate (AlGeOH) nanotubes will be initially discussed, as emerging from first-principles simulations using pure Density Functional Theory. The focus of this thesis will then change to the surface reactivity of the rutile TiO_2(110) surface by using an interplay between experiment and theory. The long standing gap between the experimentally observed and the theoretically predicted mechanisms regarding interactions of molecular oxygen and the hydroxylated TiO_2(110) surface will be bridged, by making the proposal that the oxygen vacancies on the surface are charged. This proposal will be supported by first-principles quantum mechanical modeling of the surface, supplemented by Scanning Tunneling Microscopy simulations. The agreement between the theoretical and experimental results provides confidence that this new approach will shed more light in understanding the fundamental properties regarding the reactivity of TiO_2(110). The final chapter of the thesis will discuss the electronic structure of another important surface of TiO_2, the TiO_2(011) surface. In the following two chapters of this thesis, a brief review of the theoretical and computational tools that were adopted for this analysis will be given before discussing their applications for the study of the model systems, AlSi(Ge)OH nanotubes and TiO_2 (110) and (011) surfaces.

1.2 A brief discussion on Metal Oxides

Before proceeding to the discussion of the computational techniques used in the thesis and the theoretical framework within which they are applied, it is instructive to discuss the significance of the properties of the class of materials that is the focus of subsequent chapters. Due the vast amount of literature is available on the subject only a brief high-level overview will be given to set the setting for the discussions that will follow in the subsequent chapters. An in-depth discussion of the properties of metal oxides and the knowledge that has been accumulated so far would be impossible to be given within a few pages. References to reviews and monographies on the subsect will be given in the end of this introductory chapter.

Research activity on metal oxides has significantly increased after the realisation in the 1970s that TiO_2 can facilitate the electrolysis of water without the application of a bias voltage by acting as a photocatalyst. This increase was further compounded

by the interest accumulated after the discovery in the late 1980s of metal oxide based high-temperature semiconductors. The technological applications for metal oxides also extends to catalysis. Oxides in general are used as high surface area supports of metal nanoparticles to induce new interaction channels that enhance the catalytic properties of the system. In addition, metal oxides themselves can act as catalysts in a series of technologically important chemical reactions either when in their pure form or when supporting other metal oxides. Apart from catalysis, other technological applications include photoelectrolysis and most importantly gas sensors. Interactions between the surface of the metal oxide and the gas molecules alter the electronic structure of the oxide and induce changes in its conductivity which can be measured and used to alert for the presence of the molecule of interest. Moreover, as it is known from every day experience, metals reduce to oxides when exposed to an oxidizing environment. Corrosion of metals is usually an undesirable effect that results in high costs in industrial environments and has therefore been and it is still studied extensively.

It is evident that not only the bulk form but most importantly the surfaces of metal oxides play an important role in the range of phenomena discussed briefly in the previous paragraph. As a result, there is a continuous interest in studying their properties using a combination of theoretical and experimental studies and attracting the interest of a wide range of disciplines including physics, chemistry, engineering and even biology. Despite the huge amount of crystallographic data that have been acquired since the early 20^{th} century, the structure of metal oxide surfaces is still not very well understood. The large number of atoms in the unit cell makes the identification of the location of the atoms extremely difficult. In addition, the large number of oxidation states and the range of possible interactions make the surfaces difficult to study experimentally. This is intensified even further by the complex electronic structure exhibited by the bulk and surface structures of the oxides. In addition, surface characterisation is not trivial with different experimental methods indicating contradicting properties. It is therefore difficult to produce accurate models of realistic surfaces while models based on bulk or defect-free surface systems do not account for the intricacies found on samples studied in the laboratory. Experimental techniques like Scanning Tunneling Microscopy (STM) and Atomic Force Microscopy (AFM) are now bridging the gap allowing for more accurate studies of surfaces and providing theorists with a more accurate picture for the construction of realistic models.

Within the wide range of metal oxides and their important technological applications, titanium dioxide stands out as one of the most intensively studied metal oxide system. Its range of application, the flexibility in handling it in a laboratory environment and the acquired momentum has resulted in a vast amount of literature being available. Although more details regarding the properties and electronic structure of TiO_2 is given in the corresponding chapters, it is instructive at this point to briefly discuss on a high level the applications and properties of this important metal oxide semiconductor to set the background for the rest of this thesis. As in the previous paragraphs discussing briefly metal oxides in general and due to the vast amount of publications available, references on reviews of the subject will be given at the end of this introduction. TiO_2 has important industrial applications as it is used as a gas sensor, in photocatalysis, as a coating for preventing corrosion, as a insulator in metal oxide field effective transistors (MOSFETs), in cellar cells and many more. In heterogenous catalysis since most of the catalysts consist of a metal oxide support with metal clusters supported on it, the need for understanding how the catalysis is facilitated in those systems is of high interest. In addition, the aim to identify its photoelectric and photochemical properties has been one more factor in the extensive investigations of the surface properties of TiO_2. Moreover, irradiation of TiO_2 results in electron-hole pairs traveling to the surface and interacting with adsorbates. This behaviour is exploited in wastewater purification, in coatings for self-cleaning windows, in medical applications, in thin-layers used to protect from environmental damage and many more.

The bulk structure of TiO_2 is manifested in three different variations, rutile, anatasse and brookite. Different structures can also be constructed in the laboratory under high-pressure conditions however only rutile and anatasse are usually considered for their technological applications. Defects of the bulk phase of TiO_2 play an important role in the properties of the derived surfaces with colour centres, double charged oxygen vacancies and Ti interstitials in different oxidation states dominating depending on the conditions under which the bulk structure was formed. The rutile $TiO_2(110)$ surface is the most stable and well understood surface and is therefore preferred for experiments in the laboratory. A description of the surface is given in Chapter 5. Advances in the theoretical modeling of materials along with efficient numerical schemes and faster supercomputing facilities (Chapter 5, Thijssen [1] and Pang [2], Martin [3] and Kaxiras [4], Pacheco [5] and Kirk and Hwu [6]) along with increased sophistication in experimental

techniques (e.g. Hofer and Foster [7]) has allowed for a further expanding interest in the exact physical and chemical properties in view of advancing the fundamental knowledge of the properties of the various surfaces of rutile and anatasse $TiO_2(110)$ as well as metal oxides in general and their corresponding technological applications.

A technical survey of the science behind metal oxides discussing both experimental and theoretical aspects and the bulk and surface properties can be found in Cox [8] and Henrich and Cox [9]. An extensive review of the achievements and the current state of research activity regarding TiO_2 can be found in Diebold [10].

1.3 Bibliography

[1] J. M. Thijssen. *Computational Physics.* Cambridge University Press, 1999.

[2] Tao Pang. *An Introduction to Computational Physics.* Cambridge University Press, 2010.

[3] Richard Martin. *Electronic Structure, Basic Theory and Practical Methods.* Cambridge University Press, 2004.

[4] Efthimios Kaxiras. *Atomic and Electronic Structure of Solids.* Cambridge University Press, 2003.

[5] Peter S. Pacheco. *Parallel Programming with MPI.* Morgan Kaufmann Publishers, 1997.

[6] David B. Kirk and Wen mei W. Hwu. *Programming Massively Parallel Processors.* Morgan Kaufmann Publishers, 2010.

[7] Adam S. Foster and Werner A. Hofer. *Scanning Probe Microscopy. Atomic Scale Engineering by Forces and Currents.* Springer, 2006.

[8] P. A. Cox. *Transition Metal Oxides.* Oxford University Press, 1992.

[9] V. E. Henrich and P. A. Cox. *The Surface Science of Metal Oxides.* Cambridge University Press, 1996.

[10] Ulrike Diebold. The Surface Science of Titanium Dioxide. *Surface Science Reports*, 48(5-8):53–229, 2003.

Chapter 2

Density Functional Theory

2.1 The Many-Body problem

Considering a physical system consisting of N atoms, such as a nanocluster or a surface, its low energy ground state properties can in principle be studied by relying to the non-relativistic Schrödinger equation[1]. The standard approach consists of obtaining the wavefunctions Ψ after solving Schrödinger's equation, and then of calculating all the relevant observable variables as averaged values of quantum mechanical Hermitian operators [1].

This process, which is extremely difficult even for the simplest realistic systems, can be simplified if the Born-Oppenheimer approximation is adopted [2]. Within this approximation the nuclei are considered motionless, compared to the rapidly moving electrons of the system, and give rise to a background potential in which the electrons move and interact with each other. The wavefunctions obtained this way, end up depending on the coordinates of the nuclei only parametrically, as opposed to the explicit dependence on the coordinates of the electrons themselves. The vibration, rotation and other spatial displacements of the nuclei can, in this case, be neglected. These ideas can be illustrated starting from a Hamiltonian which is rather simple in form but contains all the relevant, energetically low- lying, degrees of freedom

$$\mathcal{H} = \mathcal{H}_{\mathrm{e}} + \mathcal{H}_{\mathrm{i}} + \mathcal{H}_{\mathrm{ei}} \tag{2.1.1}$$

[1] In the case of heavier atoms where their core electrons are orbiting the nucleus with velocities approaching the speed of light, Schrödinger's equation must be replaced by Dirac's equation in order to take into account any relativistic effects.

with the terms representing the kinetic energy and interactions between the electrons, the nuclei and the interactions between the electrons and nuclei

$$\mathcal{H}_{\mathrm{e}} = -\sum_{i=1}^{M} \frac{\hbar^2}{2m}\nabla_i^2 + \frac{e^2}{4\pi\epsilon_0}\sum_{i=1}^{M}\sum_{j<i}\frac{1}{r_{ij}} \tag{2.1.2}$$

$$\mathcal{H}_{\mathrm{i}} = -\sum_{I=1}^{N} \frac{\hbar^2}{2M_I}\nabla_I^2 + \frac{e^2}{4\pi\epsilon_0}\sum_{I=1}^{N}\sum_{J<I}\frac{Z_I Z_J}{R_{IJ}} \tag{2.1.3}$$

$$\mathcal{H}_{ei} = -\frac{e^2}{4\pi\epsilon_0}\sum_{i=1}^{M}\sum_{I=1}^{N}\frac{Z_I}{r_{iI}}. \tag{2.1.4}$$

The number of atoms is denoted by N and of the electrons by M, while their positions are labeled as R_I and r_i respectively. Z_I refers to the atomic number of atom I, and $r_{iI} = \|\boldsymbol{r}_i - \boldsymbol{R}_I\|$ is the distance between the electron i positioned at $\boldsymbol{r}_i$ and the atom I at $\boldsymbol{R}_I$ (similarly for r_{ij} and R_{IJ}).

When considering the Born-Oppenheimer approximation, the kinetic terms for the nuclei are dropped and since their positions remain fixed this leads to a constant term in the Hamiltonian for the inter-nuclei Coulomb interactions. The Hamiltonian (2.1.1) can then be rewritten as

$$\mathcal{H} = -\sum_{i=1}^{M} \frac{\hbar^2}{2m}\nabla_i^2 + \frac{e^2}{4\pi\epsilon_0}\sum_{i=1}^{M}\sum_{j<i}\frac{1}{r_{ij}} + \sum_{i=1}^{M} u(\boldsymbol{r}_i) \tag{2.1.5}$$

where $u(\boldsymbol{r}_i)$ refers to the background potential created by all the nuclei in which the electrons with coordinates $\boldsymbol{r}_i$ move and is equivalent to the term (2.1.4). All terms in (2.1.3) have been dropped since the first term is zero and the second is constant leading only to a shift by a constant value in the eigenvalues of Schrödinger's equation while not affecting its eigenfunctions. Schrödinger's equation now becomes[2]

$$\mathcal{H}\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_M) = \mathcal{E}\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_M) \tag{2.1.6}$$

with the Hamiltonian $\mathcal{H}$ of (2.1.5). Ideally, one would solve (2.1.6) after defining $u(\boldsymbol{r})$ to obtain the wavefunctions Ψ before proceeding to calculating any physically relevant quantities necessary. Considering, however, that Ψ is a function of $3M$ variables, one for every spatial degree of freedom of each of the M electrons, and that the second term describing the Coulomb interaction between electrons in (2.1.5) does not allow for the decoupling of (2.1.6) into $3M$ single particle equations using standard techniques for

[2]Spin degrees of freedom have been neglected.

solving partial differential equations like separation of variables, makes this approach a rather formidable one, particularly when the number of electrons does not allow for a statistical analysis of the system[3]. Therefore, a variety of methods had to be devised, like Hartree-Fock theory [2], formulation using perturbation theory and Green's functions [3–6] and Density Functional Theory [7] among others. Density Functional theory will be the main subject of the following section.

2.2 The Thomas-Fermi Model and the Electron Gas

The first steps towards a different formulation of the problem of solving (2.1.6) can be traced back to the works of Thomas and Fermi in the 1920s. Their approach consisted of approximating the distribution of the electrons in an atom with a uniform gas of non-interacting electrons, moving under the influence of a nuclear potential at a temperature of $0K$. By dividing the space into cells containing ΔN non-interacting electrons, and by solving the ordinary Schrödinger equation for particles inside a box, the energy levels ϵ the electrons can occupy are given by

$$\epsilon = \frac{\hbar^2\pi^2}{2mL^2}(n_x^2 + n_y^2 + n_z^2) \tag{2.2.1}$$

with $n_i = 1, 2, 3, \ldots$ and $i = x, y, z$. L is the size of each side of the box. For large n_i, the number of distinct energy levels is given by

$$\Phi(\epsilon) = \frac{1}{8}\big(\frac{4}{3}\pi R^3\big) \tag{2.2.2}$$

where R is the radius of a sphere given by $R^2 = \sum_{i=x,y,z} n_i^2$. Combining the expression for R with Eq. 2.2.2 the number of energy levels is given as function of E by

$$\Phi(\epsilon) = \frac{1}{6\pi^2}\Big(\frac{2mL^2}{\hbar^2}\Big)^{3/2}\epsilon^{3/2}. \tag{2.2.3}$$

The density of the states $g(\epsilon)$ can be found from

$$\Phi(\epsilon + \delta\epsilon) = \Phi(\epsilon) + \frac{d\Phi(\epsilon)}{d\epsilon}\delta\epsilon \tag{2.2.4}$$

[3]Usually, the most difficult to study systems are those with $\mathcal{O}(10^1 - 10^2)$ degrees of freedom like nanoclusters and large molecules [3].

with $g(e) = d\Phi(\epsilon)/d\epsilon$ and $\big(\epsilon + \delta\epsilon\big)^{3/2}$ given by the McLaurin series $\epsilon^{3/2} + 3/2\epsilon^{1/2}\delta\epsilon + O(\delta\epsilon^2)$. By putting everything together, the expression for $g(\epsilon)\delta\epsilon$ becomes

$$g(\epsilon)\delta(\epsilon) = \frac{1}{4\pi^2}\Big(\frac{2mL^2}{\hbar^2}\Big)^{3/2}\epsilon^{1/2}\delta\epsilon. \tag{2.2.5}$$

Eq. 2.2.5 is the number of energy levels between ϵ and $\epsilon+\delta\epsilon$. It is therefore now possible to calculate the total energy and total number of electrons within the cell as follows. The probability distribution for states of energy ϵ to be occupied is given by the usual Fermi-Dirac distribution

$$f(\epsilon) = \frac{1}{1 + e^{\beta(\epsilon-\mu)}} \tag{2.2.6}$$

which for T = 0K reduces to the step function

$$f(\epsilon) = \begin{cases} 1 & \text{if } \epsilon \leq \epsilon_F \\ \\ 0 & \text{if } \epsilon > \epsilon_F \end{cases}. \tag{2.2.7}$$

ϵ_F is the energy of the highest occupied energy level and it is known as the Fermi energy. The total energy therefore is given by

$$\Delta E = 2\int d\epsilon\epsilon f(\epsilon)g(\epsilon) = \frac{1}{5\pi^2}\Big(\frac{2mL^2}{\hbar^2}\Big)^{3/2}\epsilon_F^{5/2} \tag{2.2.8}$$

while the total number of electrons by

$$\Delta N = 2\int d\epsilon f(\epsilon)g(\epsilon) = \frac{1}{3\pi^2}\Big(\frac{2mL^2}{\hbar^2}\Big)^{3/2}\epsilon_F^{3/2}. \tag{2.2.9}$$

Dividing the two equations, 2.2.8 and 2.2.9, together, and by eliminating the Fermi energy, the total energy can be given as a function of the total number of electrons by the following relation

$$\Delta E = (3\pi^2)^{2/3}\Big(\frac{3\hbar^2L^2}{10m}\Big)\Big(\frac{\Delta N}{L^3}\Big)^{5/3} \tag{2.2.10}$$

with

$$\frac{\Delta E}{\Delta N} = \frac{3}{5}\epsilon_F \tag{2.2.11}$$

where L^3 is the volume ΔV of the box with side L. It can therefore be seen that in the limit $\Delta V \to 0$, $\Delta N/\Delta V = n(\mathbf{r})$ is the electron density. By integrating the energy ΔE

over all space

$$T_{TF}[n(\mathbf{r})] = (3\pi^2)^{2/3}\Big(\frac{3\hbar^2 L^2}{10m}\Big)\int d\mathbf{r} n^{5/3}(\mathbf{r}) \tag{2.2.12}$$

we get the total kinetic energy of the uniform non-interacting electron gas as a functional of the charge density. The functional 2.2.12 is Thomas-Fermi kinetic energy functional which in, atomic units, reduces to

$$T_{TF}[n(\mathbf{r})] = C_F \int d\mathbf{r} n^{5/3}(\mathbf{r}), \qquad C_F = \frac{3}{10}(3\pi^2)^{2/3} = 2.871. \tag{2.2.13}$$

In Eq. 2.2.13 the explicit dependence of the charge density on $\mathbf{r}$ can be seen. The concept of this local density approximation will be discussed in the later sections.

Returning to Eq. 2.2.13, one can write a total energy functional for the uniform electron gas. This can be achieved by incorporating the Thomas-Fermi kinetic energy functional and the classical Coulomb electrostatic interactions between electrons and electrons and nuclei. This functional has the form 2.2.14, and is again dependant on the electron density only

$$E_{TF}[n(\mathbf{r})] = C_F \int d\mathbf{r} n^{5/3}(\mathbf{r}) - Z\int d\mathbf{r}\frac{n(\mathbf{r})}{r} + \frac{1}{2}\int d\mathbf{r}_1 d\mathbf{r}_2 \frac{n(\mathbf{r}_1)n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}. \tag{2.2.14}$$

The Thomas-Fermi energy functional 2.2.14 can be used to find the ground state of an atom by minimising it under the constraint

$$N[n(\mathbf{r})] = \int d\mathbf{r} n(\mathbf{r}) \tag{2.2.15}$$

using the method of Lagrange multipliers

$$\delta\left\{E_{TF}[n(\mathbf{r})] - \mu\left(\int d\mathbf{r} n(\mathbf{r}) - N\right)\right\} = 0 \tag{2.2.16}$$

with the Lagrange multiplier μ being the equilibrium chemical potential, given, after minimisation of 2.2.14, by

$$\mu = \frac{5}{3}C_F n^{2/3}(\mathbf{r}) - V(\mathbf{r}). \tag{2.2.17}$$

The potential $V(\mathbf{r})$ is the electrostatic potential

$$V(\mathbf{r}) = \frac{Z}{r} - \int d\mathbf{r}_2 \frac{n(\mathbf{r}_2)}{|\mathbf{r} - \mathbf{r}_2|} \tag{2.2.18}$$

due to the electrons and the nuclei. It is now possible to solve Eq. 2.2.17 for a given chemical potential μ and external potential $V(\mathbf{r})$, acquire the ground state charge density $n(\mathbf{r})$ and substitute it in Eq. 2.2.16 to obtain the total energy of the semi-classical uniform electron gas.

The Thomas-Fermi model was initially extended by Dirac, [8], to include the non-classical exchange energy contribution 2.2.19 to the total energy functional 2.2.14

$$K_D[n(\mathbf{r})] = C_{\mathrm{x}} \int d\mathbf{r} n^{4/3}(\mathbf{r}), \qquad \text{with} \quad C_{\mathrm{x}} = \frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} = 0.7386. \tag{2.2.19}$$

The total energy functional is then known as the Thomas-Fermi-Dirac energy functional and is given by

$$E_{TFD}[n(\mathbf{r})] = E_{TF}[n(\mathbf{r})] - C_{\mathrm{x}} \int d\mathbf{r} n^{4/3}(\mathbf{r}). \tag{2.2.20}$$

Minimisation of Eq. 2.2.20 using the method of Lagrange multipliers as in Eq. 2.2.16 results in the following relation between the chemical potential μ, the electrostatic potential $V(\mathbf{r})$ and the charge density $n(\mathbf{r})$

$$\mu = \frac{5}{3} C_F n^{2/3}(\mathbf{r}) - \frac{4}{3} C_{\mathrm{x}} n^{1/3}(\mathbf{r}) - V(\mathbf{r}) \tag{2.2.21}$$

with solutions obtained following the same scheme as in the simple Thomas-Fermi model.

It can now be seen that although the Thomas-Fermi and the Thomas-Fermi-Dirac models were a first step towards obtaining accurate quantitative results for many-body atoms and molecules, they had several obvious shortcomings that didn't make them attractive compared to other existing numerical schemes like Hartree-Fock. The two models assumed a homogeneous distribution of the electronic cloud for the computation of the kinetic and exchange energy, and therefore did not include any gradients that could account for a rapid change in the distribution of the charge. The models also failed to predict molecular bonding [9] and depressed the energies for even the simplest atoms [10]. In addition, the Thomas-Fermi model neglected quantum mechanical effects like exchange and correlation, while the addition of the Dirac term for the exchange in the total energy functional failed to improve the quantitative power of the model [7]. Despite the deficiencies of the Thomas-Fermi theory, the concept of the reformulation of the formalism around the charge density $n(\mathbf{r})$ provided the foundations on top of which Density Functional Theory was later built. The basic concepts and ideas behind DFT

are the main subjects of discussion of the following sections.

2.3 Fundamental Ideas

As it was discussed in the previous section, one of the quantities that can be calculated after solving (2.1.6) is the particle density $n(\boldsymbol{r})$

$$n(\boldsymbol{r}) = \int d\boldsymbol{r}_2 d\boldsymbol{r}_3 \dots d\boldsymbol{r}_M \Psi^*(\boldsymbol{r}, \boldsymbol{r}_2, \boldsymbol{r}_3, ..., \boldsymbol{r}_M)\Psi(\boldsymbol{r}, \boldsymbol{r}_2, \boldsymbol{r}_3, ..., \boldsymbol{r}_M). \tag{2.3.1}$$

A particle density $n(\boldsymbol{r})$ that can be written in a form as in Eq. 2.3.1 is said to be *N-representable*. This is always the case in practice [11, 12]. The difficulty in obtaining Ψ, however, has led to a different formulation, due to P. Hohenberg and W. Kohn, of the many-body problem. In this formulation, the fundamental quantity of the theory is not the ground state wavefunction Ψ_0 itself anymore, but the ground state charge density $n_0(\boldsymbol{r})$.

The key idea of this formulation is that the variables of the theory that fix the form of (2.1.5) and define Ψ through (2.1.6), namely the number of the electrons M and the potential $u(\boldsymbol{r})$, lose their status as the main variables of the theory. This is because $u(\boldsymbol{r})$ *is a unique functional of* $n(\boldsymbol{r})$*, apart from a trivial additive constant*[4][13]. Therefore, every other observable quantity that defines the properties of the ground state of the system, like the kinetic energy $T[n_0(\boldsymbol{r})]$, the potential energy $V[n_0(\boldsymbol{r})]$, and the total energy $E[n_0(\boldsymbol{r})]$, can be defined as a unique functional of the ground state charge density $n_0(\boldsymbol{r})$. Such a charge density that can be associated with an external potential $u(\boldsymbol{r})$ is said to be *u-representable*. Moreover, the number of the electrons in the system is fixed by the imposed constraint

$$M = M[n(\boldsymbol{r})] = \int d\boldsymbol{r} n(\boldsymbol{r}). \tag{2.3.2}$$

It can be shown ([7]) that the total energy of the system can now be written as

$$E_u[n(\boldsymbol{r})] = T[n(\boldsymbol{r})] + V_{ei}[n(\boldsymbol{r})] + V_e[n(\boldsymbol{r})] \tag{2.3.3}$$

where $T[n(\boldsymbol{r})]$ is the functional for the kinetic energy of the system, $V_e[n(\boldsymbol{r})] = V_c[n(\boldsymbol{r})] +$

[4] This is known as the first Hohenberg-Kohn theorem.

$V_{\mathrm{xc}}[n(\boldsymbol{r})]$ describes the classical Coulomb electron-electron interactions as well as exchange and correlation and $V_{ei}[n(\boldsymbol{r})] = \int d\boldsymbol{r} n(\boldsymbol{r})u(\boldsymbol{r})$ is the electron-ion interactions as defined by the potential $u(\boldsymbol{r})$. By putting everything together the following expression for the total energy can be derived

$$E_u[n(\boldsymbol{r})] = \int d\boldsymbol{r} n(\boldsymbol{r})u(\boldsymbol{r}) + F[n(\boldsymbol{r})] \tag{2.3.4}$$

where $F[n(\boldsymbol{r})] = T[n(\boldsymbol{r})] + V_C[n(\boldsymbol{r})] + V_{\mathrm{xc}}[n(\boldsymbol{r})]$ is a functional which is independent of the ionic configuration which gives rise to $u(\boldsymbol{r})$ as well as the number of electrons and it is therefore valid for any system. Hence it is known as a *universal functional*. By explicitly writing the classical Coulomb term in $F[n(\boldsymbol{r})]$ as

$$V_C[n(\boldsymbol{r})] = \frac{1}{2}\int d\boldsymbol{r} d\boldsymbol{r'} \frac{n(\boldsymbol{r})n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|} \tag{2.3.5}$$

the total energy of the system can be written as

$$E_u[n(\boldsymbol{r})] = \int d\boldsymbol{r} n(\boldsymbol{r})u(\boldsymbol{r}) + \frac{1}{2}\int d\boldsymbol{r} d\boldsymbol{r'} \frac{n(\boldsymbol{r})n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|} + G[n(\boldsymbol{r})] \tag{2.3.6}$$

with $G[n(\boldsymbol{r})]$ being universal too.

In principle, the charge density $n_0(\boldsymbol{r})$ that minimizes the total energy (2.3.6) needs to be found under the constraint (2.3.2) that the number of electrons of the system remain fixed. This is a variational procedure based on the fact that for an arbitrary charge density $n(\boldsymbol{r'})$ associated with an external potential $u(\boldsymbol{r'})$ the total energy $E_u[n(\boldsymbol{r'})]$ satisfies

$$E_0 \leq E_u[n(\boldsymbol{r'})] \tag{2.3.7}$$

where E_0 is the total ground state energy[5]. Minimization of (2.3.6) under the constraint (2.3.2) can proceed by employing the method of *Lagrange multipliers* in which one seeks to find the extrema, usually minima, of an auxiliary functional $\Omega[f]$ which in this case can be defined as

$$\Omega[n(\boldsymbol{r})] = E_u[n(\boldsymbol{r})] - \mu\Big\{M - \int d\boldsymbol{r} n(\boldsymbol{r})\Big\} \tag{2.3.8}$$

where μ is the Lagrange multiplier which is to be determined [6]. By varying the functional

[5] This is known as the second Hohenberg-Kohn theorem.

[6] μ is also known as the chemical potential.

(2.3.8) with respect to the charge density $n(\boldsymbol{r})$

$$\frac{\delta\Omega[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} = \frac{\delta E_u[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} - \mu\frac{\delta}{\delta n(\boldsymbol{r})}\Big\{M - \int d\boldsymbol{r} n(\boldsymbol{r})\Big\} = 0 \tag{2.3.9}$$

and using the standard rules for functional differentiation, one gets the following Euler-Lagrange equation for the ground state of the system after substituting in (2.3.9) the expression (2.3.6) for $E_u[n(\boldsymbol{r})]$:

$$\mu = u(\boldsymbol{r}) + \int d\boldsymbol{r'}\frac{n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|} + \frac{\delta G[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})}. \tag{2.3.10}$$

It can already be seen at this point that significant problems have already arisen. This is due to the lack of an exact form for the universal functional $G[n(\boldsymbol{r})]$, that can capture the physics of the quantum-mechanical exchange and correlation effects of the electrons. Although this could be a disastrous issue for the fate of Density Functional Theory, W. Kohn and L. J. Sham in their seminal 1965 work [14] provided a way to perform practical calculations within this framework. Their method will be discussed in the section below.

2.4 The Kohn-Sham Theory

Kohn's and Sham's method relies on the assumption that there can be a dual, non-physical, electron system, with the same ground state charge density as the original system. This wouldn't be much of a simplification if it wasn't for the following requirement. This new auxiliary system should consist of a non-interacting electron gas with electrons that move in an effective potential background. This, in essence, reduces the problem from a many-body to a single-particle one. The latter, in this case, is in principle exactly soluble, provided all terms in the corresponding Hamiltonian are known. Unfortunately, this is not the case and several approximations need to be made eventually in order for actual calculations to be performed.

Returning to the expression (2.3.4) for the total energy of the original many-body system

and focusing on the universal functional $F[n(\boldsymbol{r})]$, its terms can be rearranged as

$$\begin{aligned} F[n(\boldsymbol{r})] &= T[n(\boldsymbol{r})] + V_{ee}[n(\boldsymbol{r})] \\ &= T[n(\boldsymbol{r})] - T_s[n(\boldsymbol{r})] + V_{ee}[n(\boldsymbol{r})] - V_C[n(\boldsymbol{r})] + \\ &\quad + T_s[n(\boldsymbol{r})] + V_C[n(\boldsymbol{r})] \end{aligned} \tag{2.4.1}$$

where $V_{ee}[n(\boldsymbol{r})] = V_C[n(\boldsymbol{r})] + V_{\mathrm{xc}}[n(\boldsymbol{r})]$ and $T_s[n(\boldsymbol{r})]$ is an arbitrary, not known, correction to the total kinetic energy. The exact physical meaning of the term $T_s[n(\boldsymbol{r})]$ will be discussed later on. By defining the first four terms in (2.4.1) as the exchange-correlation energy of the system,

$$E_{\mathrm{xc}}[n(\boldsymbol{r})] = T[n(\boldsymbol{r})] - T_s[n(\boldsymbol{r})] + V_{ee}[n(\boldsymbol{r})] - V_C[n(\boldsymbol{r})] \tag{2.4.2}$$

the following expression for the universal functional can be obtained

$$F[n(\boldsymbol{r})] = E_{\mathrm{xc}}[n(\boldsymbol{r})] + T_s[n(\boldsymbol{r})] + V_C[n(\boldsymbol{r})]. \tag{2.4.3}$$

Substituting this expression for $F[n(\boldsymbol{r})]$ in (2.3.4) and following the same minimization procedure as in (2.3.8), (2.3.9) and (2.3.10) we get the following Euler-Lagrange equation for the ground state of the system

$$\mu = u(\boldsymbol{r}) + \int d\boldsymbol{r'} \frac{n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|} + \frac{\delta E_{\mathrm{xc}}[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} + \frac{\delta T_s[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})}. \tag{2.4.4}$$

The term $u_{\mathrm{xc}}(\boldsymbol{r}) = \delta E_{\mathrm{xc}}[n(\boldsymbol{r})]/\delta n(\boldsymbol{r})$ is identified as the exchange-correlation potential which when substituted in (2.4.4) gives

$$\begin{aligned} \mu &= u(\boldsymbol{r}) + \int d\boldsymbol{r'} \frac{n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|} + u_{\mathrm{xc}}(\boldsymbol{r}) + \frac{\delta T_s[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} \\ &= u_{\mathrm{eff}}(\boldsymbol{r}) + \frac{\delta T_s[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} \end{aligned} \tag{2.4.5}$$

with

$$u_{\mathrm{eff}}(\boldsymbol{r}) = u(\boldsymbol{r}) + \int d\boldsymbol{r'} \frac{n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|} + u_{\mathrm{xc}}(\boldsymbol{r}) \tag{2.4.6}$$

known as the Kohn-Sham effective potential. Equation (2.4.5) along with the constraint (2.3.2) corresponds to a system of M non-interacting electrons moving under the influence of the effective potential (2.4.6). One can therefore obtain the ground state charge

density of the original interacting system

$$n(\boldsymbol{r}) = \sum_{i}^{M} |\psi_i(\boldsymbol{r})|^2 \tag{2.4.7}$$

by solving the following single-particle Schrödinger-like equations

$$\Big[-\frac{\hbar^2}{2m}\nabla^2 + u_{\text{eff}}(\boldsymbol{r}) \Big]\psi_i = \epsilon_i \psi_i \tag{2.4.8}$$

using an iterative, self-consistent scheme under which

1. A random $n(\boldsymbol{r})$ is chosen.
2. Using $n(\boldsymbol{r})$, $u_{\text{eff}}(\boldsymbol{r})$ is constructed through (2.4.6).
3. With this $u_{\text{eff}}(\boldsymbol{r})$, the eigenvalue equations (2.4.8) are solved obtaining M ϵ_i's and ψ_i's.
4. A new charge density $n_{\text{new}}(\boldsymbol{r})$ is computed using (2.4.7).
5. The process is repeated with the new charge density $n_{\text{new}}(\boldsymbol{r})$ substituted in step 1 until $n(\boldsymbol{r})$ converges.

In practice, some mixing scheme is used that mixes the old with the new charge density before repeating the iteration. This is in order to avoid instabilities in the calculation [15, 16]. Although the charge densities $n(\boldsymbol{r})$ participating in this iterative scheme have to be u-representable, this is not guaranteed by default. One can expect however that this requirement will be satisfied if the initial charge density is a relatively good guess of the physical one [17]. After the self-consistency cycle has ended, and a converged value for the charge density is obtained then the observables, like the total energy of the system can be calculated.

In addition, one is also interested in finding the most energetically low-lying geometry of the system under study. This can be achieved by embedding the above iterative scheme within a loop under which after the end of each electronic cycle, the forces exerted on atoms are calculated using the Feynman-Hellmann theorem. Within the Kohn-Sham independent-particle formulation, the expression for the forces can take the form [18]

$$\mathbf{F}_I = -2\text{Re}\sum_i \int d\mathbf{r} \frac{\partial \psi_i^*}{\partial \mathbf{R}_I}\left[\frac{1}{2}\nabla^2 + V_{\text{KS}} - \epsilon_i\right]\psi_i - \int d\mathbf{r}\left[V_{\text{KS}} - V^{\text{in}}\right]\frac{\partial n}{\partial \mathbf{R}_I} \tag{2.4.9}$$

where the first term in 2.4.9 is exactly zero when a plane-wave basis set is employed, due to its independence from the positions of the atoms. This is not the case however in implementations where atomic orbitals, Gaussians or other position dependent functions are used as basis sets, mostly because they do not span the entire space completely. In this case, corrections to forces need to be included or different theoretical schemes employed [19]. The atoms are then displaced towards the direction that minimizes the forces, usually no more than a predefined distance, and then the self-consistent scheme is repeated for the new geometry. The calculation ends when all spatial components of the forces are less than a chosen cut-off which in modern electronic structure calculations can be significantly less[7] than 0.1 eV/Å.

Despite the apparent robustness of the above scheme, an exact expression for the u_{xc} term in 2.4.6, or alternatively for the $E_{\mathrm{xc}}[n(\boldsymbol{r})]$ in 2.4.4, which holds for all of the materials is still not known. It therefore needs to be approximated in a case by case basis. Exactly because of this , there has been a significant amount of effort in the development of accurate exchange-correlation functionals. This has allowed for research-quality calculations to increasingly more sophisticated materials including semiconductors and insulators. A brief review of those functionals that were relevant to our work will be given in the following section.

2.5 Derivation of the Kohn-Sham equations

Consider the expression of the total energy of the interacting system Eq. 2.3.4

$$E_u[n(\boldsymbol{r})] = \int d\boldsymbol{r} n(\boldsymbol{r}) u(\boldsymbol{r}) + F[n(\boldsymbol{r})] \tag{2.5.1}$$

with

$$F[n(\boldsymbol{r})] = E_{\mathrm{xc}}[n(\boldsymbol{r})] + T_s[n(\boldsymbol{r})] + V_C[n(\boldsymbol{r})]. \tag{2.5.2}$$

as in Eq. 2.4.3. In the absence of any interactions i.e. for a non-interacting electron gas, Eq. 2.5.2 reduces to

$$F[n(\boldsymbol{r})] = T_s[n(\boldsymbol{r})] \tag{2.5.3}$$

[7] Up to an order of magnitude.

giving for Eq. 2.5.1

$$E_{u_s}[n_s(\boldsymbol{r})] = \int d\boldsymbol{r} n_s(\boldsymbol{r}) u_s(\boldsymbol{r}) + T_s[n_s(\boldsymbol{r})] \tag{2.5.4}$$

where $n_s(\boldsymbol{r})$ is the charge density of the non-interacting system. By observing the equivalent equations for the interacting case in sections 2.3 and 2.4 it can be seen that in order for $n_s(\boldsymbol{r})$ to equal the charge density of the fully interacting system $n(\boldsymbol{r})$, the non-interacting electrons need to move under the influence of the potential

$$u_s(\boldsymbol{r}) \equiv u_{\text{eff}}(\boldsymbol{r}) = u(\boldsymbol{r}) + \int d\boldsymbol{r'} \frac{n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|} + u_{\text{xc}}(\boldsymbol{r}). \tag{2.5.5}$$

The dynamics of the system are then described by the Schrödinger-like single-particle Kohn-Sham equations

$$\Big[-\frac{\hbar^2}{2m}\nabla^2 + u_{\text{eff}}(\boldsymbol{r}) \Big] \psi_i = \epsilon_i \psi_i. \tag{2.5.6}$$

The role of $T_s[n(\boldsymbol{r})]$ has also been elucidated. It corresponds to the kinetic energy of the non-interacting non-physical auxiliary electrons propagating under the influence of the effective potential Eq. 2.5.5, with wavefunctions derived from the solution of the Kohn-Sham equations Eq. 2.5.6. The value of the kinetic energy is

$$T_s[n(\boldsymbol{r})] = \sum_{i=1}^{N} \langle \psi_i(\boldsymbol{r}) | -\frac{\hbar^2}{2m}\nabla^2 | \psi_i(\boldsymbol{r}) \rangle \tag{2.5.7}$$

while the total charge density for both the interacting and the non-interacting system is

$$n(\boldsymbol{r}) = \sum_{i=1}^{N} \sum_{j=-1/2}^{1/2} |\psi_{i,j}(\boldsymbol{r})|^2 \tag{2.5.8}$$

where N is the total number of electrons in the system.

2.6 Exchange and Correlation

The pursuit for exchange-correlation functionals that yield accurate results for a wide range of materials, has been a long one. Since Density Functional Theory became a practical tool for electronic structure calculations, the demand for increased accuracy combined with the increased performance of modern computers led to a systematic investigation of approximations to $E_{\text{xc}}[n(\boldsymbol{r})]$ as appearing in 2.4.4. Knowledge of $E_{\text{xc}}[n(\boldsymbol{r})]$

could enable the study of the chemistry of atoms, molecules and solids in their ground state[8]. Before proceeding to the different exchange- correlation functionals, the notion of the exchange-correlation hole will be discussed due to its importance in the development of accurate approximations to exchange and correlation.

An exchange-correlation hole is in essence a depletion, known as a hole, of electronic charge that surrounds an electron due to quantum mechanical interactions such as non-classical Coulomb interactions and the imposed Fermi-Dirac statistics [7] which has the effect of reducing the probability of finding an electron in the vicinity of another. Effectively, the exchange-correlation energy can be thought of as the energy due to the interaction between the electron and its exchange-correlation hole

$$E_{\mathrm{xc}}[n] = \frac{1}{2}\int d\mathbf{r}d\mathbf{r}' \frac{n(\mathbf{r})n_{\mathrm{xc}}(\mathbf{r},\mathbf{r}'-\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} \tag{2.6.1}$$

where

$$n_{\mathrm{xc}}(\mathbf{r},\mathbf{r}'-\mathbf{r}) = n(\mathbf{r}')\int_0^1 d\lambda(g(\mathbf{r},\mathbf{r}',\lambda)-1). \tag{2.6.2}$$

The variable λ in Eq. 2.6.2 emerges from the so-called adiabatic connection in which the non-interacting system 2.4.8 is related to the physical one through the Hamiltonian [21]

$$H_\lambda = -\frac{1}{2}\nabla^2 + V_{\mathrm{ext}}(\mathbf{r}) + V_\lambda + \lambda V_{ee}. \tag{2.6.3}$$

For $\lambda = 0$ the non-interacting system is obtained while the fully interacting physical system corresponds to $\lambda = 1$. These two limiting cases are connected through a continuum of partially interacting systems all of which share the same physical charge density of the fully interacting system. $g(\mathbf{r},\mathbf{r}',\lambda)-1)$ in Eq. 2.6.2 is the pair-correlation function of the system 2.6.3 with density $n(\mathbf{r})$ and Coulomb interaction λV_{ee} [7, 18, 22]. The contribution to the exchange-correlation hole n_{xc} can be divided to an exchange and a correlation part, $n_{\mathrm{xc}} = n_{\mathrm{x}} + n_{\mathrm{c}}$, with the two components satisfying the following

[8] Time-dependent DFT, [20], that allows for excited states will not be discussed here.

constraints which are known as sum rules [23]

$$n_{\mathrm{x}}(\mathbf{r},\mathbf{r}) = -\frac{n(\mathbf{r})}{2} \tag{2.6.4}$$

$$n_{\mathrm{x}}(\mathbf{r},\mathbf{r}'-\mathbf{r}) \leq 0 \tag{2.6.5}$$

$$\int d\mathbf{r}' n_{\mathrm{x}}(\mathbf{r},\mathbf{r}'-\mathbf{r}) = -1 \tag{2.6.6}$$

$$\int d\mathbf{r}' n_{\mathrm{c}}(\mathbf{r},\mathbf{r}'-\mathbf{r}) = 0 \tag{2.6.7}$$

and therefore

$$\int d\mathbf{r}' n_{\mathrm{xc}}(\mathbf{r},\mathbf{r}'-\mathbf{r}) = -1 \tag{2.6.8}$$

for the entire exchange-correlation hole. This implies that charge depletion of an exchange-correlation hole is exactly one electron. The compliance to the constraints 2.6.4- 2.6.8 has been an indication of the success of certain approaches to approximations to the exchange-correlation energy Eq. 2.6.1.

The first attempt to approximate $E_{\mathrm{xc}}[n(\boldsymbol{r})]$ was based on the Thomas-Fermi theory for the homogeneous electron gas which was discussed in section 2.2. For this system, the kinetic energy of the gas per volume is given by

$$t_{\mathrm{hom}} = \frac{3\hbar^2}{10m}(3\pi^2)^{2/3} n^{5/3} \tag{2.6.9}$$

where m is the electron mass and n the electron density. Since in reality the density can fluctuate, it is possible to rewrite Eq. 2.6.9 to account for this by introducing a dependence of the density on the spatial position $\mathbf{r}$ assuming that the density varies only weakly and slowly with space

$$t_{\mathrm{hom}}[n(\mathbf{r})] = \frac{3\hbar^2}{10m}(3\pi^2)^{2/3} n^{5/3}(\mathbf{r}). \tag{2.6.10}$$

It is now possible to get an expression for the kinetic energy within the approximation $n \to n(\mathbf{r})$ by integrating Eq. 2.6.10 over all real space

$$T^{\mathrm{LDA}}[n(\mathbf{r})] = \frac{3\hbar^2}{10m}(3\pi^2)^{2/3} \int d\mathbf{r} n^{5/3}(\mathbf{r}). \tag{2.6.11}$$

The approximation $n = n(\mathbf{r})$ in Eq. 2.6.11 was the first attempt to use a non-constant, spatially depended, expression for the charge density within the Thomas-Fermi model.

This approach was extended by introducing the Local Density Approximation (LDA) within which the exchange-correlation energy is defined as

$$E_{\mathrm{xc}}^{\mathrm{TF}}[n(\mathbf{r})] = \int n(\mathbf{r})\epsilon_{\mathrm{xc}}[n(\mathbf{r})]d\mathbf{r} \tag{2.6.12}$$

with the exchange-correlation energy density $\epsilon_{\mathrm{xc}}[n(\mathbf{r})]$ decomposing to two parts

$$\epsilon_{\mathrm{xc}}[n(\mathbf{r})] = \epsilon_{\mathrm{x}}[n(\mathbf{r})] + \epsilon_{\mathrm{c}}[n(\mathbf{r})] \tag{2.6.13}$$

the first corresponding to exchange effects while the second to correlation. By taking the functional derivative of Eq. 2.6.12 with respect to the density $n(\mathbf{r})$, the exchange-correlation potential appearing in the effective potential Eq. 2.4.6 of the Kohn-Sham equation Eq. 2.4.8 can be calculated as

$$u_{\mathrm{xc}}^{\mathrm{LDA}}(\mathbf{r}) = \epsilon_{xc}[n(\mathbf{r})] + n(\mathbf{r})\frac{\partial\epsilon_{\mathrm{xc}}[n(\mathbf{r})]}{\partial n(\mathbf{r})} \tag{2.6.14}$$

The LDA method has been incorporated in all DFT implementations for the approximation of $E_{\mathrm{xc}}[n(\boldsymbol{r})]$. This has been achieved by relying on the exact expression for the exchange energy per particle of a homogeneous electron gas which is given by

$$\epsilon_{\mathrm{x}}^{\mathrm{LDA}}[n(\mathbf{r})] = -\frac{3q^2}{4}\left(\frac{3}{\pi}\right)^{1/3} n^{1/3}(\mathbf{r}). \tag{2.6.15}$$

Using the same rationale that led from Eq. 2.6.9 to Eq. 2.6.11, an expression for the total exchange energy can be obtained

$$E_{\mathrm{x}}^{\mathrm{LDA}}[n(\mathbf{r})] = -\frac{3q^2}{4}\left(\frac{3}{\pi}\right)\int d\mathbf{r} n^{4/3}(\mathbf{r}). \tag{2.6.16}$$

Now it remains for the $E_{\mathrm{c}}^{\mathrm{LDA}}[n(\mathbf{r})]$ part of the exchange-correlation energy to be approximated. An exact expression for the correlation energy however is not known and this is the point where approximations enter the LDA framework. There have been several approaches to the subject and each one improved on the previous attempts, starting from perturbation theory and the random phase approximation (RPA), [24–27], to stochastic methods like the quantum Monte-Carlo calculations performed by Ceperley and Alder [28] and improved by Perdew, Zunger, Wang and others [29–31]. It should be noted

that although the expression

$$E_{\mathrm{xc}}[n(\mathbf{r})] = \int d\mathbf{r} n(\mathbf{r}) \epsilon_{\mathrm{xc}}[n(\mathbf{r})] \tag{2.6.17}$$

with $\epsilon_{\mathrm{xc}}[n(\mathbf{r})] = \epsilon_{\mathrm{x}} n[(\mathbf{r})] + \epsilon_{\mathrm{c}} n[(\mathbf{r})]$ does not account for spin, it can be generalized in the case where spin polarization is present to

$$E_{\mathrm{xc}}[n_{1/2}(\mathbf{r}), n_{-1/2}(\mathbf{r})] = \int d\mathbf{r} n(\mathbf{r}) \epsilon_{\mathrm{xc}}[n_{1/2}(\mathbf{r}), n_{-1/2}(\mathbf{r})] \tag{2.6.18}$$

with $n(\mathbf{r}) = n_{1/2}(\mathbf{r}) + n_{-1/2}(\mathbf{r})$ where $n_{1/2}(\mathbf{r}), n_{-1/2}(\mathbf{r})$ are the spin densities of the homogeneous electron gas for electrons with spin up and down respectively.

Despite the success of L(S)DA, the increased accuracy needed for quantum chemistry calculations has prompted the need for further improvement. The inefficiencies of L(S)DA lie on the fact that Eq. 2.6.17 and 2.6.18 are valid only near the limit of a slowly varying electron density. Although this is a reasonable assumption for a homogeneous electron gas, it does not hold very well in the case of real atoms, molecules and solids. The next logical step was to consider those variations in the density by incorporating density gradients that account for the rate of the spatial variation of $n(\mathbf{r})$.

Non-local forms for $E_{\mathrm{xc}}[n(\boldsymbol{r})]$ were initially introduced as corrections to Eq. 2.6.18 within a scheme known as the Gradient Expansion Approximation (GEA) [32–34], e.g

$$E_{\mathrm{xc}} = E_{\mathrm{xc}}^{\mathrm{LDA}} + \int d\mathbf{r} B_{\mathrm{xc}}[n(\mathbf{r})] |\nabla n(\mathbf{r})|^2 \tag{2.6.19}$$

where B_{xc} was written as

$$B_{\mathrm{xc}} = \frac{e^2}{n^{4/3}(\mathbf{r})} C_{\mathrm{xc}}(r_s), \quad r_s = \left[\frac{1}{4\pi n(\mathbf{r})}\right]^{1/3} \tag{2.6.20}$$

with $C_{\mathrm{xc}}(r_s)$ a quantity obtained using perturbation theory within the Random Phase Approximation (RPA) and parametrically fitted to already available data obtained by experiment or other more accurate but expensive calculations [33, 35, 36]. Although GEA appeared to be a systematic improvement over L(S)DA mathematically wise, it emerged that results reported within this framework tended to be worse than those obtained with L(S)DA [25, 37–39] because the corrections introduced violated certain physical constraints and sum rules Eq. 2.6.4-2.6.8 in which L(S)DA fared better mostly

due to systematic cancelations between errors [23, 30, 40, 41].

The failure of GEA resulted in the emergence of the idea of representing the exchange and correlation energy by a functional of arbitrary dependence on $n(\mathbf{r})$ and $\nabla n(\mathbf{r})$, a scheme referred to as the Generalized Gradient Approximation (GGA)

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[n(\mathbf{r})] = \int d\mathbf{r} f[n(\mathbf{r}), \nabla n(\mathbf{r})]. \tag{2.6.21}$$

There have been suggested many expressions for f, other based on physical arguments and constraints and other, more empirical ones, on data sets fitted in analytic expressions for f, or combinations of both however the first steps towards that direction involved an expression for f similar to GEA 2.6.19 with the addition of additional imposed constraints that eliminated the shortcomings of GEA. Perdew and Wang suggested that one could improve upon GEA by introducing a cut-off radius R_{c} to the GEA exchange-hole in order to satisfy the exchange hole sum rules [42]. They proposed the following form, usually referred to as PW86, for the exchange energy

$$E_{\mathrm{x}}^{\mathrm{GGA}}[n(\mathbf{r})] = A_{\mathrm{x}} \int d\mathbf{r} n^{4/3}(\mathbf{r}) F^{\mathbf{GGA}}(s) \tag{2.6.22}$$

with $A_{\mathrm{x}} = -3/4(3/\pi)^{1/3}$ and

$$F^{\mathrm{GGA}}(s) = \left(1 + 0.0864\frac{s^2}{m} + bs^4 + cs^6\right)^m \tag{2.6.23}$$

where $m = 1/15$, $b = 14$, $c = 0.2$ and $s = |\nabla n(\mathbf{r})|/(2k_F n(\mathbf{r}))$ and with the spin-density functional generalization of Eq. 2.6.22

$$E_{\mathrm{x}}^{\mathrm{GGA}}[n_{1/2}(\mathbf{r}), n_{-1/2}(\mathbf{r})] = \frac{1}{2}E_{\mathrm{x}}^{\mathrm{GGA}}[n_{1/2}(\mathbf{r})] + \frac{1}{2}E_{\mathrm{x}}^{\mathrm{GGA}}[n_{-1/2}(\mathbf{r})]. \tag{2.6.24}$$

Perdew and Wang further improved upon PW86 in their PW91 exchange-correlation functional [43] by incorporating corrections suggested by Becke [44] for the correct asymptotic behaviour of the exchange energy density and Johnson, Gill and Pople who studied the behaviour of a range of functionals and compared them with other ab-initio techniques like Hartree-Fock and Moller-Plesset [45]. In their work, they suggested the

revised form for $F(s)$ in Eq. 2.6.22

$$F(s) = \frac{1 + 0.19645s\sinh^{-1}(7.7956s) + (0.2743 - 0.1508e^{-100s^2})s^2}{1 + 0.19645s\sinh^{-1}(7.7956s) + 0.004s^4}. \tag{2.6.25}$$

Their expression for the correlation energy part of the functional was given by

$$E_{\rm c}[n_{1/2}(\mathbf{r}), n_{-1/2}(\mathbf{r})] = \int d\mathbf{r} n[\epsilon_{\rm c}(r_s, \zeta) + H(t, r_s, \zeta)] \tag{2.6.26}$$

with

$$t = \frac{|\nabla n(\mathbf{r})|}{2gk_s n(\mathbf{r})} \tag{2.6.27}$$

$$g = \frac{[(1+\zeta)^{2/3} + (1-\zeta)^{2/3}]}{2} \tag{2.6.28}$$

$$k_s = \left(\frac{4k_F}{\pi}\right)^{1/2} \tag{2.6.29}$$

$$\zeta = \frac{n_{1/2}(\mathbf{r}) - n_{-1/2}(\mathbf{r})}{n_{1/2}(\mathbf{r}) + n_{-1/2}(\mathbf{r})}. \tag{2.6.30}$$

H was defined as

$$\begin{aligned} H = {} & g^3 \frac{\beta^2}{2\alpha} \ln\left[1 + \frac{2\alpha}{\beta}\frac{t^2 + At^4}{1 + At^2 + A^2t^4}\right] + \\ & + \nu\left[C_{\rm c}(r_s) - C_{\rm c}(0) - \frac{3}{7}C_{\rm x}\right] \times g^3 t^2 \exp\left[-100g^4\left(\frac{k_s}{k_F}\right)^2 t^2\right] \end{aligned} \tag{2.6.31}$$

with

$$A = \frac{2\alpha}{\beta}\frac{1}{e^{-2\alpha\epsilon_{\rm c}(r_s,\zeta)/(g^3\beta^2)} - 1} \tag{2.6.32}$$

and $\alpha = 0.09$, $\beta = \nu C_{\rm c}(0)$, $\nu = (16/\pi)(3\pi^2)^{1/3}$, $C_{\rm c}(0) = 0.004235$ and $C_{\rm x} = -0.001667$. Expressions for $\epsilon_{\rm c}(r_s, \zeta)$ which included self- interaction corrections to the energy were given by Perdew and Zunger [29] and for $C_{\rm c}(r_s)$ by Rasolt and Geldart [33].

It is quite apparent that the PW91 exchange-correlation functional is a rather complicated one. This prompted Perdew, Burke and Ernzerhof to search for a rather simplified version that would retain all the good features of it. In their 1996 work [46] they derived the following expressions for exchange and correlation

$$E_C^{\rm GGA}[n_{1/2}(\mathbf{r}), n_{-1/2}(\mathbf{r})] = \int d\mathbf{r} n[\epsilon_C^{\rm unif}(r_s, \zeta) + H(r_s, \zeta, t)] \tag{2.6.33}$$

with

$$t = \frac{|\nabla n(\mathbf{r})|}{2\phi k_s n(\mathbf{r})} \tag{2.6.34}$$

$$k_s = \left(\frac{4k_F}{\pi}\right)^{1/2} \tag{2.6.35}$$

$$\zeta = \frac{n_{1/2}(\mathbf{r}) - n_{-1/2}(\mathbf{r})}{n_{1/2}(\mathbf{r}) + n_{-1/2}(\mathbf{r})} \tag{2.6.36}$$

$$\phi(\zeta) = \frac{1}{2}[(1+\zeta)^{2/3} + (1-\zeta)^{2/3}] \tag{2.6.37}$$

were they showed that

$$H = \frac{e^2}{\alpha_0}\gamma\phi^3 \ln\left[1 + \frac{\beta}{\gamma}t^2\left(\frac{1+At^2}{1+At^2+A^2t^4}\right)\right] \tag{2.6.38}$$

with

$$A = \frac{\beta}{\gamma}\left[\exp\left(-\frac{\epsilon_C^{\text{unif}}}{\gamma\phi^3 e^2/\alpha_0}\right) - 1\right]^{-1} \tag{2.6.39}$$

has the correct behaviour under the rapidly and slowly varying and the high-density limit i.e

$$t \to \infty, \qquad H \to -\epsilon_C^{\text{unif}} \tag{2.6.40}$$

$$t \to 0, \qquad H \to \beta\phi^3 t^2 \tag{2.6.41}$$

$$n(\mathbf{r}) \to \lambda^3 n(\lambda\mathbf{r}) \quad \text{and} \quad \lambda \to \infty, \qquad H \to (\text{const}). \tag{2.6.42}$$

For they exchange they showed that

$$E_X[n(\mathbf{r})] = \int d\mathbf{r} n \epsilon_{\text{x}}^{\text{unif}}(n) F_X(s) \tag{2.6.43}$$

with

$$F_X(s) = 1 + k - \frac{k}{[1 + \frac{\mu s^2}{k}]} \tag{2.6.44}$$

where $\epsilon_{\text{x}}^{\text{unif}}(n) = -3k_f/4\pi$, $k_F = (3\pi^2 n)^{1/3}$, $k = 0.804$ and $s = |\nabla n|/2k_F n$. Zhang and Yang argued that a value $k = 1.245$ improves even further total and atomization energies [47]. The PBE exchange- correlation functional was found to yield similar or better results to PW91 while being simpler and therefore easier to improve upon having retained the correct features of the L(S)DA even after introducing non-local terms and density gradients unlike the previous GEA approaches [46, 48]. A different approach was

adopted by Lee, Young and Parr [49] who transformed the expression for the correlation energy obtained by Colle and Salvetti [50] using Hartree-Fock theory and fittings to data sets obtained for the Helium atom to a density functional. Their expression for the correlation energy, later simplified by Miechlich et. al. [51], was

$$E_{\mathrm{c}}[n(\mathbf{r})] = -\alpha \int d\mathbf{r} \frac{1}{1+dn(\mathbf{r})^{-1/3}} \Bigg\{ n(\mathbf{r}) + \\ + bn^{-2/3}(\mathbf{r}) \left[C_F n^{5/3}(\mathbf{r}) - 2t_W + \left(\frac{1}{9} t_W + \frac{1}{18} \nabla^2 n(\mathbf{r}) \right) e^{-cn^{-1/3}(\mathbf{r})} \right] \Bigg\} \quad (2.6.45)$$

with $C_F = \frac{3}{10}(3\pi^2)^{2/3}$ and

$$t_W = \frac{1}{8} \frac{|\nabla n(\mathbf{r})|^2}{n(\mathbf{r})} - \frac{1}{8} \nabla^2 n(\mathbf{r}). \quad (2.6.46)$$

The combination of Becke's expression for exchange [52] with Eq. 2.6.45 for correlation is known as the BLYP functional. GGA implementations like BLYP, PW91 and PBE provide an overally better accuracy compared to L(S)DA particularly in cases where inhomogeneities in the charge density are dominant and improve the bond lengths and angles as well as energies [46, 48, 52–57] and they have therefore been implemented in all major electronic structure codes. They are used routinely for the studies of properties of a wide range of materials.

2.7 Modern Approaches in Exchange and Correlation

Despite the success of various implementations of GGA, not all issues were addressed leaving room for further improvement. It was argued that much like L(S)DA, GGA did not solve the problem of spurious electron self-interactions that have been noticed since the early days of the Thomas-Fermi theory [58]. As it can be seen in the expression for the Coulomb functional Eq. 2.3.5, Coulomb interactions do not vanish even when there is only one electron present. Perdew and Zunger introduced self-interaction corrections to L(S) DA [29] however Goedecker and Umrigar pointed out that L(S)DA and GGA approximations performed better [56]. Becke suggested that since for a non-interacting system with $\lambda = 0$ (see Eq. 2.6.2) the exchange-correlation energy 2.6.1 is in principle void of any correlation and therefore corresponds to exact exchange only, a mixing of exact exchange in the exchange-correlation functional E_{xc} would account for the lack

of correlation and could be the way forward in this respect [59, 60]. He proposed the following form for exchange and correlation

$$E_{\mathrm{xc}} = E_{\mathrm{xc}}^{\mathrm{L(S)DA}} + a_0(E_{\mathrm{x}}^{\mathrm{exact}} - E_{\mathrm{x}}^{\mathrm{L(S)DA}}) + a_{\mathrm{x}}\Delta E_{\mathrm{x}}^{\mathrm{B88}} + a_{\mathrm{c}}\Delta E_{\mathrm{c}}^{\mathrm{PW91}} \tag{2.7.1}$$

where $\Delta E_{\mathrm{c}}^{\mathrm{PW91}}$ is the gradient correction to correlation due to Perdew and Wang [43] and $\Delta E_{\mathrm{x}}^{\mathrm{B88}}$ Becke's correction to exchange [44]. The approximation Eq. 2.7.1 is a three-parameter semi-empirical one with the values for a_0, a_{x} and a_{c} fitted to experimental data. A more recent reincarnation of this approximation has been the B3LYP functional in which Becke's correction is combined with the correction to correlation suggested by Lee, Yang and Parr [49]. The problem with this approach, is that these functionals work really well on the class of materials on which the empirical parameters were fitted, however from a physicist's perspective it would be more satisfactory to reduce the number of free parameters and/or fix the functional under certain physical constraints. This has the advantage of offering a better understanding on how and why the approximation does or doesn't work, it can be used with materials outside the fitted data set, and can be improved by physical reasoning. Becke suggested an alternative form to Eq. 2.7.1 with only one parameter

$$E_{\mathrm{xc}} = E_{\mathrm{xc}}^{\mathrm{GGA}} + a_0(E_{\mathrm{x}}^{\mathrm{exact}} - E_{\mathrm{x}}^{\mathrm{GGA}}) \tag{2.7.2}$$

and Perdew, Ernzerhof and Burke fixed the value of a_0 to $1/4$ using perturbation theory [53, 61]. Following that line of thought, Ernzerhof and Scuseria [62] simultaneously with Adamo and Barone [63] suggested a mixing of the PBE functional with exact exchange, the PBE benefiting from having no free parameters, with the form, known as PBE0,

$$E_{\mathrm{xc}} = aE_{\mathrm{x}}^{\mathrm{HF}} + (1-a)E_{\mathrm{x}}^{\mathrm{PBE}} + E_{\mathrm{c}}^{\mathrm{PBE}}. \tag{2.7.3}$$

Their results showed a very good agreement between Eq. 2.7.3 and other semi-empirical functionals. In their efforts to produce an accurate approximation to exchange and correlation, without relying to empirical data and fitting parameters, that would work for metallic systems, Heyd, Scuseria and Ernzerhof, based their ideas on employing a Coulomb potential for the exchange interaction between electrons. This potential should have the effect of screening the long-range part of the Hartree-Fock exchange [64]. This

was achieved by separating the long-range from the short-range terms in the Coulomb operator which was the cause of the misbehaviour in metals due to the singularity present at $k = 0$ in the Fourier transformation of $1/r$. By employing the PBE0 exchange-correlation functional Eq. 2.7.3 and splitting the exchange energy into short-range and long-range terms they obtained the following approximation for exchange and correlation

$$E_{\mathrm{xc}} = aE_{\mathrm{x}}^{\mathrm{HF,SR}}(\omega) + (1-a)E_{\mathrm{x}}^{\mathrm{PBE,SR}}(\omega) + E_{\mathrm{x}}^{\mathrm{PBE,LR}} + E_{\mathrm{c}}^{\mathrm{PBE}} \tag{2.7.4}$$

where ω is an adjustable parameter stemming from the Coulomb operator separation

$$\frac{1}{r} = \frac{1-\mathrm{erf}\omega}{r} + \frac{\mathrm{erf}\omega}{r}. \tag{2.7.5}$$

The approximation Eq. 2.7.4 has since then been incorporated into schemes using Gaussians as well as plane-waves basis sets. It was found that it gave substantially accurate results for semiconductors, with a modest increase in computer time compared to pure DFT methods like L(S)DA and less computational cost compared to other hybrid implementations like PBE0 [64–68].

A different approach can be employed for strongly correlated systems for which the standard pure-DFT methods fail spectacularly. For many simple transition metal-oxides, for which localised $d-$ and $f-$ electrons play an important role, pure-DFT cannot predict, for example, their insulating behaviour. It severely underestimates the band gap and even predicts a metallic ground state [69]. Based on the Hubbard model for strongly correlated systems, L(S)DA+U [70, 71] attempts to address this issue. In this framework, the exchange-correlation functional is modified to read

$$E_{\mathrm{LDA+U}}[n(\mathbf{r})] = E_{\mathrm{LDA}}[n(\mathbf{r})] + E_{\mathrm{Hub}}[\{n_m^{I\sigma}\}] - E_{\mathrm{DC}}[\{n^{I\sigma}\}] \tag{2.7.6}$$

where $n_m^{I\sigma}$ is the occupation number for atom I, E_{LDA} is the exchange-correlation functional within the LDA approximation, E_{Hub} is the Hubbard term and E_{DC} is included

to compensate for the double counting [70]. The Hubbard term is given by

$$H_{\mathrm{Hub}}[\{n^I_{mm'}\}] = \frac{1}{2}\sum_{\{m\},\sigma,I} \langle m,m''|V_{ee}|m'm''\rangle\, n^{I\sigma}_{mm'} n^{I\sigma}_{m'm''} + \frac{1}{2}\sum_{\{m\},\sigma,I} \left(\langle m,m''|V_{ee}|m',m'''\rangle - \langle m,m''|V_{ee}|m''',m'\rangle\right) n^{I\sigma}_{mm'} n^{I\sigma}_{m'm''} \tag{2.7.7}$$

with

$$\langle m,m''|V_{ee}|m',m''\rangle = \sum_{k=0}^{2l} a_k(m,m',m'',m''')F^k. \tag{2.7.8}$$

F^k are the radial Slater integrals and l is the angular momentum of the localised $d-$ and $f-$ electrons. a_k is given by

$$a_k(m,m',m'',m''') = \frac{4\pi}{2k+1}\sum_{q=-k}^{k} \langle lm|Y_{kq}|lm'\rangle \langle lm''|Y^*_{kq|lm'''}\rangle\,. \tag{2.7.9}$$

In addition, the double-counting compensation term in Eq. 2.7.7 is given by

$$E_{\mathrm{DC}}[\{n^I\}] = \sum_I \frac{U}{2} n^I(n^I-1) - \sum_I \frac{J}{2}[n^{I\uparrow}(n^{I\uparrow}-1) + n^{I\downarrow}(n^{I\downarrow}-1)] \tag{2.7.10}$$

with $n^I = n^{I\uparrow} + n^{I\downarrow}$. For d electrons[9] the parameters U and J describing the screened on-site Coulomb interaction are given, in terms of the Slater orbitals F^K, by

$$U = \frac{1}{(2l+1)^2}\sum_{m,m'} \langle m,m'|V_{ee}|m,m'\rangle = \frac{1}{(2l+1)^2}\sum_{m,m'} U_{mm'} = F^0 \tag{2.7.11}$$

$$J = \frac{1}{2l(2l+1)}\sum_{m\neq m',m'} \langle m,m'|V_{ee}|m',m\rangle = \frac{1}{2l(2l+1)}\sum_{m\neq m',m'} J_{mm'} = \frac{F^2+F^4}{14}. \tag{2.7.12}$$

The total energy functional Eq. 2.7.6 can then be written in terms of U, J and the total number of electrons as [71]

$$E_{\mathrm{LDA+U}}[n(\mathbf{r})] = E_{\mathrm{LDA}}[n(\mathbf{r})] - \left[\frac{1}{2}UN(N-1) - \frac{1}{4}JN(N-2)\right] + \frac{1}{2}\sum_{m,m',\sigma} U_{m,m'} n_{m\sigma} n_{m'-\sigma} + \frac{1}{4}\sum_{m\neq m',m',\sigma} (U_{m,m'} - J_{m,m'}) n_{m\sigma} n_{m'\sigma}. \tag{2.7.13}$$

[9] Like the Ti(3d) states discussed in chapters 5 and 6.

By differentiation of Eq. 2.7.13 with the occupation number $n_{m\sigma}$, the one-electron potential is obtained

$$\begin{aligned} V_{m\sigma}(\mathbf{r}) =& V_{\mathrm{LDA}}(\mathbf{r}) + \sum_{m'}(U_{mm'} - U_{\mathrm{eff}})n_{m-\sigma} \\ &+ \sum_{m'\neq m}(U_{mm'} - J_{mm'} - U_{\mathrm{eff}})n_{m\sigma} \\ &+ U_{\mathrm{eff}}(\frac{1}{2} - n_{m\sigma}) - \frac{1}{4}J \end{aligned} \tag{2.7.14}$$

with $U_{\mathrm{eff}} = U - J/2$. Equations 2.7.6-2.7.14 describe the general L(S)DA+U framework [70–72] that has, in addition, been enhanced and extended recently [69] and applied to a range of materials including titanium dioxide [73].

2.8 Plane-wave DFT

In making DFT practical, one has to agree upon certain conventions of which perhaps the most decisive one is the choice of the basis set. A few aspects of practical issues when a plane-wave basis set is chosen will be discussed in this section. This choice is dictated by the popularity of plane-wave based DFT codes within the physics community and also by our own adoption of the scheme.

The effect of the periodicity of a solid can be reflected on its wavefunctions by virtue of Bloch's theorem. The electronic wavefunction can be factorized into a periodic and a wave-like part

$$\psi_i(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} f_i(\mathbf{r}) \tag{2.8.1}$$

with $f_i(\mathbf{r})$ having the same periodicity as the solid [74]. By using a complete basis set of plane-waves with wavevectors the infinite set of the reciprocal lattice vectors of the crystal $\{e^{i\mathbf{G}\cdot\mathbf{r}}\}$ with $\mathbf{G}\cdot\mathbf{T} = 2\pi n$ and $n = 1, 2, 3, \ldots$, the periodic function $f_i(\mathbf{r})$ can be expanded as an infinite sum of basis functions

$$f_i(\mathbf{r}) = \sum_{\mathbf{G}} c_{i,\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \tag{2.8.2}$$

which when substituted in Eq. 2.8.1 gives

$$\psi_i(\mathbf{r}) = \sum_{\mathbf{G}} c_{i,\mathbf{G}+\mathbf{k}} e^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{r}}. \tag{2.8.3}$$

This has the advantage of expanding a wavefunction as a sum of infinite but discrete plane-waves which therefore makes it possible to choose a convenient number of waves to be included in the expansion by selecting a predefined energy cut-off $\frac{1}{2}|\mathbf{k}+\mathbf{G}_c|^2$. Although this unavoidably introduces errors in the calculation, they can be eliminated by increasing in a step by step process the cut-off energy until the total energy converges. Then the value for the cut-off can stay fixed for the remaining part of the calculation. Such a systematic increase of the basis set by altering only one parameter, the cut-off energy, is a distinct feature of plane-wave methods and it is not found in other schemes which employ different basis sets like Gaussians or localized atomic orbitals.

After introducing the wavefunction in the form Eq. 2.8.3, the Kohn-Sham equations can be reformulated by substituting Eq. 2.8.3 to Eq. 2.4.8. The outcome after integration over $\mathbf{r}$ is

$$\sum_{\mathbf{G}} c_{i,\mathbf{G}+\mathbf{k}} \left[\frac{1}{2}|\mathbf{k}+\mathbf{G}|^2\delta_{\mathbf{G}\mathbf{G}'} + u_{\text{eff}}(\mathbf{G}-\mathbf{G}')\right] = \epsilon_i \sum_{\mathbf{G}} c_{i,\mathbf{G}} \tag{2.8.4}$$

where

$$u_{\text{eff}}(\mathbf{G}-\mathbf{G}') = \int d\mathbf{r} u_{\text{eff}}(\mathbf{r}) e^{i(\mathbf{G}-\mathbf{G}'+\mathbf{k})\cdot\mathbf{r}} \tag{2.8.5}$$

is the Fourier transformation of the effective potential Eq. 2.4.6. The solution of Eq. 2.8.4 can proceed by diagonalization of the expression in the brackets in Eq. 2.8.4. Although direct diagonalization seems to be the standard approach, usually more computationally efficient techniques are employed which involve minimization of the exchange-correlation functional using steepest descents and conjugate gradients [75, 76]. As it can be seen, the size of the matrix to be diagonalized depends on the number of plane waves and therefore on the energy cut-off. As it was discussed earlier, a systematic improvement on the accuracy of the calculation can be achieved by just increasing the energy cut-off. There remains however the issue of the large number of plane waves needed to describe the rapidly oscillating wavefunctions near the core region of the atom. This problem can be tackled by using the pseudopotential method which will be discussed in the next section.

In principle, after introducing the cut-off energy, it remains for the wavefunction to be evaluated over the wavevectors $\mathbf{k}$. Due to the periodicity, the infinite extend of the bulk solid and the Bloch theorem, an infinite amount of $\mathbf{k}$-points are needed for the evaluation of the wavefunction. Then one is presented with the situation where collective properties

need to be evaluated usually in the form of integrals

$$I = \int_{\mathrm{BZ}} d\mathbf{k} f(\mathbf{k}) = \frac{(2\pi)^3}{V} \bar{f} \tag{2.8.6}$$

over the Brillouin[10] zone of the crystal, with $\bar{f}$ being the average value of f in the region of integration. Baldereschi argued that by exploiting the crystal's symmetry one can find a special $\mathbf{k}$-point, the mean-value point $\mathbf{k}^*$, for which

$$\bar{f} = f(\mathbf{k}^*) = \alpha_0 \tag{2.8.7}$$

with α_0 evaluated by expanding f to a series

$$f(\mathbf{k}) = \sum_0^\infty a_i G_i(\mathbf{k}), \qquad G_0(\mathbf{k}) = 1 \tag{2.8.8}$$

and the imposed constraint $G_i(\mathbf{k})$ for $i = 1, 2, \ldots, n$ with n the largest integer possible and Eq. 2.8.8 rapidly convergent [77].

A more general and computationally efficient approach which eventually became the standard in BZ sampling in modern electronic structure codes is due to Monkhorst and Pack [78]. By defining the vector

$$\mathbf{k}_{prs} = u_p \mathbf{b}_1 + u_r \mathbf{b}_2 + u_s \mathbf{b}_3 \tag{2.8.9}$$

with

$$u_r = \frac{2r - q - 1}{2q}, \qquad r = 1, 2, 3, \ldots, q \tag{2.8.10}$$

where q determines the number of special $\mathbf{k}$-points, q^3, and $\mathbf{b}_i$ the primitive reciprocal lattice vectors

$$\mathbf{b}_1 = \frac{2\pi}{V} \mathbf{a}_2 \times \mathbf{a}_3, \quad \mathbf{b}_2 = \frac{2\pi}{V} \mathbf{a}_3 \times \mathbf{a}_1, \quad \mathbf{b}_1 = \frac{2\pi}{V} \mathbf{a}_1 \times \mathbf{a}_2, \tag{2.8.11}$$

[10] The irreducible Brillouin zone is defined as the Wigner-Seitz cell of the reciprocal lattice.

with $\mathbf{a}_i$ the primitive lattice vectors, and the quantities

$$A_m(\mathbf{k}) = N_m^{-1/2} \sum_{|\mathbf{R}|=c_m} e^{i\mathbf{k}\cdot\mathbf{R}} \tag{2.8.12}$$

$$S_{mn}(q) = \frac{1}{q^3} \sum_{p,r,s=1}^{q} A_m^*(\mathbf{k}_{prs}) A_n(\mathbf{k}_{prs}) \tag{2.8.13}$$

with $\mathbf{R}$ all those lattice vectors $\mathbf{R} = R_1\mathbf{a}_1 + R_2\mathbf{a}_2 + R_3\mathbf{a}_3$, also known as a star, related with each other by the point symmetry of the solid, they were able to approximate the function f by the series

$$f(\mathbf{k}) = \sum_{m=1}^{\infty} f_m A_m(\mathbf{k}). \tag{2.8.14}$$

By exploiting the point-group symmetry of the solid, Eq. 2.8.13 can be re-written as

$$S_{mn}(q) = \frac{1}{q^3} \sum_{j=1}^{P(q)} w_j A_m(\mathbf{k}_j) A_n(\mathbf{k}_j) \tag{2.8.15}$$

where $\mathbf{k}_j$ are those symmetry-related $\mathbf{k}$- points in the set $\mathbf{k}_{prs}$ that lie inside the irreducible part of the Brillouin zone and w_j is a weight factor. Returning to Eq. 2.8.12 and 2.8.13, they showed that by imposing the constraint $|R_i| < q/2$ with $i = 1, 2, 3$ the coordinates of the lattice vector $\mathbf{R}$, $S_{mn}(q) = \delta_{mn}$ therefore guaranteeing the orthogonality of A_m's. Because of the orthogonality of A_m, f_m can be written as

$$f_m = \frac{V}{8\pi^3} \int_{\mathrm{BZ}} d\mathbf{k} A_m^*(\mathbf{k}) f(\mathbf{k}) \tag{2.8.16}$$

and therefore the integral Eq. 2.8.6 can be written as

$$\int_{\mathrm{BZ}} d\mathbf{k} f(\mathbf{k}) = \frac{(2\pi)^3}{V} f_1. \tag{2.8.17}$$

The function Eq. 2.8.16 can then be approximated by $\tilde{f}_m$

$$\tilde{f}_m = \frac{1}{q^3} \sum_{j=1}^{P(q)} w_j f(\mathbf{k}_j) A_m(\mathbf{k}_j) \tag{2.8.18}$$

with A_m all the functions 2.8.12 constrained by $|R_i| < q/2$. By adopting the Monkhorst-Pack scheme one can reduce the $\mathbf{k}$-point sampling to those special $\mathbf{k}$-points that are within the irreducible Brillouin zone making the evaluation of functions and integrals more efficient computationally.

2.9 Pseudopotentials

One of the main disadvantages of the plane-wave implementation of DFT is the large number of plane-waves needed to describe the rapidly varying wavefunctions of the core electrons of the atoms, thereby increasing tremendously the computational cost needed to evaluate and perform electronic structure calculations. The pseudopotential scheme is based on the idea that core electrons are chemical inert and therefore they can be substituted as long as their effects can be accurately accounted for. This is usually done by utilizing a pseudopotential that is able to reproduce the all-electron behaviour of the electronic states outside a specified region. Although various schemes of pseudopotential generation have been proposed, a few basic characteristics will be discussed in this section that underline the fundamental qualities a pseudopotential must posses.

The starting point of a pseudopotential generation scheme is the solution of the radial Kohn-Sham equation for the given isolated all-electron atom

$$\Big\{ -\frac{1}{2}\frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + V[\rho;r]\Big\} rR_{nl}^{\mathrm{AE}}(r) = \epsilon_{nl} rR_{nl}^{\mathrm{AE}}(r) \tag{2.9.1}$$

with $V[\rho;r] = -\frac{Z}{r} + V_H[\rho;r] + V_{\mathrm{xc}}^{\mathrm{LDA}}[\rho(r)]$ and $R_{nl}^{\mathrm{AE}}(r)$ the radial part of the all-electron wavefunction. After obtaining the radial wavefunctions, the pseudopotential is generated by inverting Eq. 2.9.1

$$V_{\mathrm{scr},l}^{\mathrm{PP}}(r) = \epsilon_{nl} - \frac{l(l+1)}{2r^2} + \frac{1}{2rR_{nl}^{\mathrm{PP}}}\frac{d^2}{dr^2}\Big[rR_{nl}^{\mathrm{PP}}\Big]. \tag{2.9.2}$$

In order for Eq. 2.9.2 to qualify for a pseudopotential, it must satisfy the following conditions:

1. It is better for the wavefunctions generated by the pseudopotential to be as smooth as possible. This is desirable because less plane waves are then needed for their representation.

2. The pseudo-wavefunction associated with the angular momentum l should match the all-electron one beyond a chosen, l-dependent, cut-off radius r_{cl}

$$R_l^{\mathrm{PP}}(r) = R_l^{\mathrm{AE}}, \qquad r > r_{\mathrm{cl}}. \tag{2.9.3}$$

3. Pseudo-wavefunctions should satisfy the norm-conservation condition

$$\int_0^{r_{cl}} drr^2 |R_l^{\mathrm{PP}}(r)|^2 = \int_0^{r_{cl}} drr^2 |R_l^{\mathrm{AE}}(r)|^2 \tag{2.9.4}$$

   which states that the charge inside the core region is equal for both wavefunctions.

4. Energy eigenvalues for both all-electron and pseudo-potentials must be equal

$$\epsilon_l^{\mathrm{PP}} = \epsilon_l^{\mathrm{PP}}. \tag{2.9.5}$$

By imposing the additional constraint that $R_l^{\mathrm{PP}}(r) \sim r^l$ near the origin i.e for $r \to 0$, $V_{\mathrm{scr},l}^{\mathrm{PP}}(r) = \epsilon_{nl}$ and the singularities associated with the $-Z/r$ all-electron potential are hence eliminated.

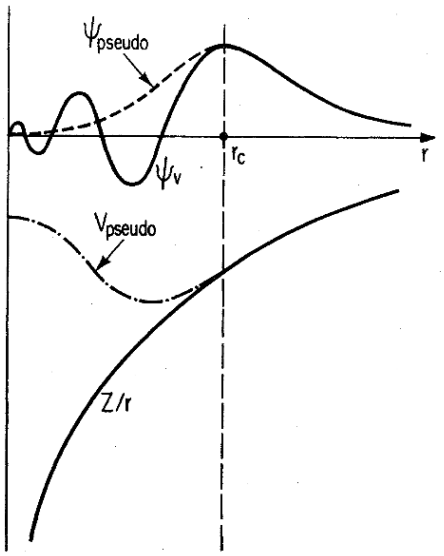

Figure 2.1: Illustration of a pseudopotential and a pseudowavefunction in comparison with the all-electron case [75].

Although several techniques have been developed that allow for increased transferability of the pseudopotential, it is possible to retain the all-electron wavefunction, and therefore improve the quality of the calculation, by adopting the Projector Augmented Wave (PAW) method [79]. The PAW method effectively replaces the rapidly varying valence wavefunctions by decomposing it to a smoothly varying part and a core part that corresponds to the region near the nucleus. The core wavefunctions are still accounted for by using the pseudopotential formalism. The breakthrough of the formalism is that the entire set of the all-electron valence wavefunctions is retained. The central aspect of the

PAW formalism is the transformation

$$\psi = T\tilde{\psi}, \qquad T = I + T_0 \tag{2.9.6}$$

of the, rapidly oscillating near the nucleus, one-particle valence wavefunctions ψ, with $\tilde{\psi}$ being a set of smooth auxiliary functions. T_0 in Eq. 2.9.6 has the form

$$T_0 = I + \sum_R S_R \tag{2.9.7}$$

where $\mathbf{R}$ defines non-overlapping spheres centered around the atoms. By expanding the real and auxiliary wavefunctions as linear combinations of plane waves within the spheres $\mathbf{R}$, the real wavefunction can be written as [18, 79]

$$|\psi\rangle = |\tilde{\psi}\rangle + \sum_m c_m\{|\psi_m\rangle - |\tilde{\psi}_m\rangle\}. \tag{2.9.8}$$

By introducing projector functions $|p_m\rangle$ localised inside the spheres $\mathbf{R}$ i.e with $p(\mathbf{r}) = 0, \quad \mathbf{r} < \mathbf{R}$, the real all-electron valence wavefunctions in Eq. 2.9.8 along with the transformation T in Eq. 2.9.6 can be written as [18, 79]

$$|\psi\rangle = |\tilde{\psi}\rangle + \sum_m \Big(|\psi_m\rangle - |\tilde{\psi}_m\rangle \Big) \langle p_m|\tilde{\psi}\rangle \tag{2.9.9}$$

$$T = I + \sum_m \Big(|\psi_m\rangle - |\tilde{\psi_m}\rangle \Big) \langle p_m| \,. \tag{2.9.10}$$

By applying the formalism in Eq. 2.9.9, expectation values of operators can be evaluated as follows

$$\langle A\rangle = \sum_m c_m \langle\psi_m|A|\psi_m\rangle = \sum_m c_m \langle\tilde{\psi}_m|T^\dagger AT|\tilde{\psi}_m\rangle \,. \tag{2.9.11}$$

As an example, the particle density $n(\mathbf{r})$ can be written as

$$n(\mathbf{r}) = \tilde{n}(\mathbf{r}) + \sum_R \Big(n^1_R(\mathbf{r}) - \tilde{n}^1_R(\mathbf{r}) \Big) \tag{2.9.12}$$

with $\tilde{n}$ being the smooth part which can be evaluated on a plane wave grid and n^1 and $\tilde{n}^1$ evaluated on a radial grid centered around the atoms described by the augmentation spheres $\mathbf{R}$. A similar expression can be obtained for the energy

$$E = \tilde{E} + \sum_R \Big(E^1_R - \tilde{E}^1_R \Big). \tag{2.9.13}$$

Although there are many subtleties in the PAW formalism that are not discussed in this section, PAW improves upon the pseudopotential method regarding the transferability of the pseudopotentials, and provides a more accurate method for calculating electronic structure properties of materials by virtue of the inclusion of the all-electron valence wavefunctions. PAW has been used routinely in our studies of the reactivity of the Rutile $TiO_2(110)$ surface and the electronic structure of the $TiO_2(011)$ surface that will be discussed in following sections.

Density Functional Theory has matured substantially since the early days of Thomas, Fermi and Dirac and the breakthrough works of Hohenberg, Kohn and Sham. It is now routinely applied not only to atoms and molecules but to more complex systems as well. There exists a variety of electronic structure codes offering different basis sets, from plane waves to localised atomic orbitals and Gaussians, each with its own advantages and disadvantages. Different pseudopotentials are also available, from the easy to understand and implement pure LDA to more sophisticated hybrid schemes. Most of the codes range from proprietary to completely open source and are highly parallelised with the ability to scale to thousands of computing nodes. Relatively recent breakthroughs include variations of the theory, that will not be discussed in this section, to allow for a transition from the expected $O\left(N^3\right)$ scaling of the standard implementations to $O\left(N\right)$, with N the number of atoms in the simulated cell. Linear scaling has allowed for the application of ab-initio, or first principles, methods to massive systems with up to 10^6 atoms. The possibility of using quantum mechanical modeling to study the electronic structure of large macro-molecules along with the associated chemical and physical processes that take place in these scales, open new horizons in other disciplines from earth sciences to biology. The discussion of Density Functional Theory presented in this chapter is not, by any means, exhaustive. Many appealing aspects including various numerical schemes have been left out due to the vast amount of literature available. It does provide, however, an outline of the theory behind the computations performed while studying the electronic structure of the systems that are discussed in subsequent chapters.

2.10 Bibliography

[1] Stephen Gasiorowicz. *Quantum Physics*. Wiley, 2003.

[2] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry, Introduction to*

Advanced Electronic Structure Theory. Dover Publications, Inc., 1989.

[3] Alexander Atland and Ben Simons. *Condensed Matter Field Theory*. Cambridge University Press, 2007.

[4] Henrik Bruus and Karsten Flensberg. *Many-Body Quantum Theory in Condensed Matter Physics*. Oxford University Press, 2004.

[5] Jorgen Rammer. *Quantum Field Theory of Non-Equilibrium States*. Cambridge University Press, 2007.

[6] Xiao-Gang Wen. *Quantum Field Theory of Many-Body Systems*. Oxford University Press, 2004.

[7] Robert G. Parr and Weitao Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press, 1989.

[8] P A M Dirac. Note on Exchange Phenomena in the Thomas Atom. *Mathematical Proceedings of the Cambridge Philosophical Society*, 26:376, 1930.

[9] Edward Teller. On the Stability of Molecules in the Thomas-Fermi Theory. *Review of Modern Physics*, 34:627, 1962.

[10] Robert G Parr and Swapan K Ghosh. Thomas-Fermi theory for Atomic Systems. *Proceedings of the National Academy of Sciences of the United States of America*, 83(1):3577, 1986.

[11] T. L. Gilbert. Hohenberg-Kohn Theorem for Nonlocal External Potentials. *Physical Review B*, 12(6):2111–2120, 1975.

[12] John E. Harriman. Orthonormal Orbitals for the Representation of an Arbitrary Density. *Physical Review A*, 24(2):680–682, 1981.

[13] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Physical Review*, 136(3B):B864–B871, 1964.

[14] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140(4A):A1133–A1138, 1965.

[15] C. G. Broyden. A Class of Methods for Solving Nonlinear Simultaneous Equations. *Mathematics of Computation*, 19(92):557–593, 1965.

[16] D. D. Johnson. Modified Broyden's Method for Accelerating Convergence in Self-Consistent Calculations. *Physical Review B*, 38(18):12807–12813, 1988.

[17] W. Kohn. v-Representability and Density Functional Theory. *Physical Review Letters*, 51(17):1596–1598, 1983.

[18] Richard Martin. *Electronic Structure, Basic Theory and Practical Methods*. Cambridge University Press, 2004.

[19] P. Pulay. Ab Initio Calculation of Force Constants and Equilibrium Geometries in Polyatomic Molecules. I. Theory. *Molecular Physics*, 17(2):197–204, 2002.

[20] M. Petersilka, U. J. Gossmann, and E. K. U. Gross. Excitation Energies from Time-Dependent Density-Functional Theory. *Physical Review Letters*, 76(8):1212–1215, 1996.

[21] J. Harris and R. O. Jones. The Surface Energy of a Bounded Electron Gas. *Journal of Physics F: Metal Physics*, 4(8):1170, 1974.

[22] Efthimios Kaxiras. *Atomic and Electronic Structure of Solids*. Cambridge University Press, 2003.

[23] John P. Perdew. Accurate Density Functional for the Energy: Real-Space Cutoff of the Gradient Expansion for the Exchange Hole. *Physical Review Letters*, 55(16):1665–1668, 1985.

[24] U. von Barth and L. Hedin. A Local Exchange-Correlation Potential for the Spin Polarized Case. *Journal of Physics C: Solid State Physics*, 5(13):1629, 1972.

[25] O. Gunnarsson and B. I. Lundqvist. Exchange and Correlation in Atoms, Molecules, and Solids by the Spin-Density-Functional Formalism. *Physical Review B*, 13(10):4274–4298, 1976.

[26] David C. Langreth and M. J. Mehl. Easily Implementable Nonlocal Exchange-Correlation Energy Functional. *Physical Review Letters*, 47(6):446–450, 1981.

[27] David C. Langreth and M. J. Mehl. Beyond the Local-Density Approximation in Calculations of Ground-State Electronic Properties. *Physical Review B*, 28(4):1809–1834, 1982.

[28] D. M. Ceperley and B. J. Alder. Ground State of the Electron Gas by a Stochastic Method. *Physical Review Letters*, 45(7):556–569, 1980.

[29] J. P. Perdew and Alex Zunger. Self-Interaction Correction to Density-Functional Approximations for Many-Electron-Systems. *Physical Review B*, 23(10):5048–5079, 1981.

[30] John P. Perdew and Yue Wang. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. *Physical Review B*, 45(23):13244–13249, 1992.

[31] S. H. Vosko, L. Wilk, and M. Nusair. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: a Critical Analysis. *Canadian Journal of Physics*, 58(8):1200–1211, 1980.

[32] David C. Langreth and S. H. Vosko. Exact Electron-Gas Response Functions at High Density. *Physical Review Letters*, 59(4):497–500, 1987.

[33] Mark Rasolt and D. J. W. Geldart. Exchange and Correlation Energy in a Nonuniform Fermion Fluid. *Physical Review B*, 34(2):1325–1328, 1986.

[34] Leonard Kleinman and Seongbok Lee. Gradient Expansion of the Exchange-Energy Density Functional: Effect of Taking Limits in the Wrong Order. *Physical Review B*, 37(9):4634–4636, 1988.

[35] Mark Rasolt and D. J. W. Geldart. Gradient Corrections in the Exchange and Correlation Energy of an Inhomogeneous Electron Gas. *Physical Review Letters*, 35(18):1234–1237, 1975.

[36] Mark Rasolt and D. J. W. Geldart. Exchange and Correlation Energy of an Inhomogeneous Electron Gas at Metallic Densities. *Physical Review B*, 13(4):1477–1488, 1976.

[37] John P. Perdew. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Physical Review B*, 33(12):8822–8824, 1986.

[38] John P. Perdew and David C. Langreth. Corrections to the Local Density Approximation: Gradient Expansion versus Wave-Vector Analysis for the Metallic Surface Problem. *Physical Review Letters*, 38(18):1030–1033, 1977.

[39] Shang-Keng Ma and Keith A. Brueckner. Correlation Energy of an Electron Gas with a Slowly Varying High Density. *Physical Review*, 165:18–31, 1968.

[40] David C. Langreth and John P. Perdew. Theory of Nonuniform Electronic Sys-

tems. I. Analysis of the Gradient Approximation and a Generalization that Works. *Physical Review B*, 21(12):5469–5493, 1980.

[41] Yue Wang et al. Exchange Potentials in Density-Functional Theory. *Physical Review A*, 41(1):78–86, 1990.

[42] John P. Perdew and Yue Wang. Accurate and Simple Density Functional for the Electronic Exchange Energy: Generalized Gradient Approximation. *Physical Review B*, 33(12):8800–8802, 1986.

[43] John P. Perdew et al. Atoms, Molecules, Solids and Surfaces: Applications of the Generalized Gradient Approximation for Exchange and Correlation. *Physical Review B*, 46(11):6671–6687, 1992.

[44] A. D. Becke. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Physical Review A*, 38(6):3098–30100, 1988.

[45] Benny G. Johnson, Peter M. W. Gill, and John A. Pople. The Performance of a Family of Density Functional Methods. *Journal of Chemical Physics*, 98(7):5612, 1993.

[46] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Physical Review Letters*, 77(18):3865–3868, 1996.

[47] Yingkai Zhang and Weitao Yang. Comment on Generalized Gradient Approximation Made Simple. *Physical Review Letters*, 80(4):890, 1998.

[48] John P. Perdew, Kieron Burke, and Yue Wang. Generalized Gradient Approximation for the Exchange-Correlation Hole of a Many-Electron System. *Physical Review B*, 54(23):16533–16539, 1996.

[49] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti Correlation-Energy Formula Into a Functional of the Electron Density. *Physical Review B*, 37(2):785–789, 1988.

[50] Renato Colle and Oriano Salvetti. Approximate Calculation of the Correlation Energy for the Closed Shells. *Theoretical Chemistry Accounts: Theory, Computation and Modelling*, 37(4):329–334, 1975.

[51] Burkhard Miechlich, Andreas Savin, Hermann Stoll, and Heinzwerner Preuss. Re-

sults Obtained with the Correlation Energy Density Functionals of Becke and Lee, Yang and Parr. *Chemical Physics Letters*, 157(3):200–206, 1989.

[52] A. D. Becke. Density-Functional Thermochemistry. I. The Effect of the Exchange-Only Gradient Correction. *Journal of Chemical Physics*, 96(3):2155, 1992.

[53] A. D. Becke. Density-Functional Thermochemistry. II. The Effect of the Perdew-Wang Generalized-Gradient Correlation Correction. *Journal of Chemical Physics*, 97(12):9173, 1992.

[54] Andrea Dal Corso, Alfredo Pasquarello, Alfonso Baldereschi, and Roberto Car. Generalized-Gradient Approximations to Density-Functional Theory: A Comparative Study for Atoms and Solids. *Physical Review B*, 53(3):1180–1185, 1996.

[55] B. Hammer, L. B. Hansen, and J. K. Norskov. Improved Adsorption Energetics Within Density-Functional Theory Using Revised Perdew-Burke-Ernzerhof Functionals. *Physical Review B*, 53(11):7413–7421, 1999.

[56] S. Goedecker and C. J. Umrigar. Critical Assessment of the Self-Interaction-Corrected Local-Density-Functional Method and its Algorithmic Implementation. *Physical Review A*, 55(3):1765–1771, 1997.

[57] Joachim Paier, Robin Hirschl, Martijn Marsman, and Georg Kresse. The Perdew-Burke-Ernzerhof Exchange-Correlation Functional Applied to the G2-1 Test Set Using a Plane-Wave Basis Set. *The Journal of Chemical Physics*, 122(23):234102, 2005.

[58] E. Fermi and E. Amaldi. Le Orbite Oos Degli Elementi. *Accad. Ital. Rome*, 1934.

[59] Axel D. Becke. Density Functional Thermochemistry III. The Role of Exact Exchange. *Journal of Physical Chemistry*, 98(7):5648–5652, 1993.

[60] Axel D. Becke. Density Functional Thermochemistry. IV. A New Dynamical Correlation Functional and Implications for Exact Exchange Mixing. *Journal of Chemical Physics*, 104(3):1040, 1996.

[61] John P. Perdew, Matthias Ernzerhof, and Kieron Burke. Rationale for Mixing Exact Exchange with Density Functional Approximations. *Journal of Chemical Physics*, 105(22):9982, 1996.

[62] Matthias Ernzerhof and Gustavo E. Scuseria. Assessment of the Perdew-

Burke-Ernzerhof Exchange-Correlation Functional. *Journal of Chemical Physics*, 110(11):5029, 1998.

[63] Carlo Adamo and Vincenzo Barone. Towards Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *Journal of Chemical Physics*, 110(13):6158, 1999.

[64] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid Functionals Based on a Screened Coulomb Potential. *Journal of Chemical Physics*, 118(18):8207, 2003.

[65] Jochen Heyd, Juan E. Peralta, Gustavo E. Scuseria, and Richard L. Martin. Energy Band Gaps and Lattice Parameters Evaluated with the Heyd-Scuseria-Ernzerhof Screened Hybrid Functional. *The Journal of Chemical Physics*, 123(17):174101, 2005.

[66] J. Paier et al. Screened Hybrid Density Functionals Applied to Solids. *The Journal of Chemical Physics*, 124(15):154709, 2006.

[67] Juan E. Peralta, Jochen Heyd, Gustavo E. Scuseria, and Richard L. Martin. Spin-Orbit Splittings and Energy Band Gaps Calculated with the Heyd-Scuseria-Ernzerhof Screened Hybrid Functional. *Physical Review B*, 74(7):073101, 2006.

[68] Jochen Heyd and Gustavo E. Scuseria. Efficient Hybrid Density Functional Calculations in Solids: Assessment of the Heyd-Scuseria-Ernzerhof Screened Coulomb Hybrid Functional. *Journal of Chemical Physics*, 121(3):1187, 2004.

[69] Matteo Cococcioni and Stefano de Gironcoli. Linear Response Approach to the Calculation of the Effective Interaction Parameters in the LDA+U Method. *Physical Review B*, 71:035105, 2005.

[70] Vladimir I. Anisimov, Jan Zaanen, and Ole K. Andersen. Band Theory and Mott Insulators: Hubbard U Instead of Stoner I. *Physical Review B*, 44(3):943–954, 1991.

[71] V. I. Anisimov, I. V. Solovyev, and M. A. Korotin. Density-Functional Theory and NiO Photoemission Spectra. *Physical Review B*, 48(23):16929–16934, 1993.

[72] Vladimir I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein. First-Principles Calculations of the Electronic Structure and Spectra of Strongly Correlated Systems: The LDA+U Method. *Journal of Physics: Condensed Matter*, 9:767, 1997.

[73] CJ Calzado, NC Hernandez, and JF Sanz. Effect of On-Site Coulomb Repulsion term U on the Band-Gap States of the Reduced Rutile (110)TiO2 Surface. *Physical Review B*, 77(4):045118, 2008.

[74] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Thomson Learning, 1976.

[75] M. C. Payne et al.˙ Iterative Minimization Techniques for Ab-Initio Total Energy Calculations: Molecular Dynamics and Conjugate Gradients. *Reviews of Modern Physics*, 64(4):1045–1097, 1992.

[76] I. Stich and R. Car and M. Parrinello and S. Baroni. Conjugate Gradient Minimization of the Energy Functional: A New Method for Electronic Structure Calculation. *Physical Review B*, 39(8):4997–5004, 1989.

[77] A. Baldereschi. Mean-Value Point in the Brillouin Zone. *Physical Review B*, 7(12):5212–5215, 1973.

[78] Hendrik J. Monkhorst and James D. Pack. Special Points for Brillouin Zone Integrations. *Physical Review B*, 13(12):5188–5192, 1976.

[79] P. E. Blochl. Projector Augmented Wave Method. *Physical Review B*, 50(24):17953–17979, 1994.

Chapter 3

Scanning Tunneling Microscopy

3.1 Introduction

The Scanning Tunneling Microscope (STM) is an experimental technique based on the quantum mechanical tunneling of electrons. It was originally developed in the late 1970s and early 1980s by Binnig and Rohrer, for real-space atomic resolution microscopy of surfaces [1–3]. The experimental apparatus, Fig. 3.1, consists of a probe, a sharp conducting scanning tip suspended above the surface of interest. The tip is mounted at the edge of a piezoelectric tube which allows for the lateral and vertical movement of the tip, by application of voltage on its electrodes. The tip is connected through a current amplifier to a feedback loop which moves the tip depending on the tunneling current recorded when a bias voltage is applied between the tip and the surface. Although in principle the ambient environment for STM experiments can vary, the entire set-up is usually submerged inside a vacuum chamber to avoid contamination of the surface. Due to the very small separations between the tip and the surface, the STM is very prone to environmental vibrations and therefore it has to be suspended on vibrational absorber to eliminate noise [4].

The fundamental idea behind Binnig and Rohrer's work was that by applying a bias voltage between the tip and the surface, tunneling electrons begin to flow between the two and by scanning the surface while keeping the tunneling current constant, which usually done automatically by changing the tip-surface separation through the feedback mechanism, one can obtain a constant-current contour map of the surface which can reveal more information about the surface's structure [2, 5]. The invention of the STM

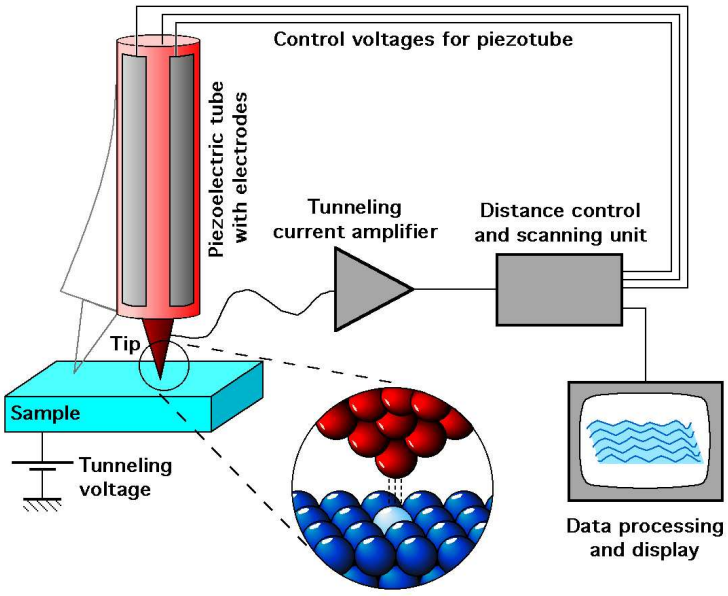

Figure 3.1: Schematic view of the STM. Figure: Michael Schmid, TU Wien.

has given a unique capability for studying surfaces with atomic resolution. Although it was initially used for the study of the reconstruction of the Si(110) 7×7 surface [2], experiments are now systematically been performed in a wide range of materials. These studies are done with the aim to elucidate the chemical properties of surfaces like reconstructions and relaxations, reaction mechanisms like adsorption and dissociation of atoms and molecules, electronic structure, surface defects, atom manipulation and many more [6].

Despite the widespread use of the STM however, the large number of factors that can influence the experimentally measured quantities accompanied by a rather complicated theoretical model of how the STM actually works, can make the experimental results relatively difficult to understand and interpret. There has, therefore, been established a trend of collaboration between experiment and theory in order to shed light to the mechanisms that prevail during the STM operation, correct and improve theoretical models and explain experimental findings. A brief review of the theory behind the Scanning Tunneling Microscope will be given in the following sections.

3.2 Theoretical Formulation

There have been several approaches to formulating a theoretical model that can accurately quantify the tunneling current between the tip of the STM and the surface under

study. The, perhaps, most instructive one is derived from Bardeen's approach to tunneling in metal-insulator-metal junctions [7]. The tip and the surface can be considered as two infinite leads, acting as electron reservoirs, separated by a vacuum region. In this case, the following Hamiltonian can be written for the coupled system

$$\mathcal{H} = \mathcal{H}_T + \mathcal{H}_S + V = \mathcal{H}_0 + V \tag{3.2.1}$$

with V a weak perturbation describing the tunneling of electrons between the weakly coupled tip and surface. $\mathcal{H}_T$, $\mathcal{H}_S$ is the independent Hamiltonians for the tip and the surface respectively

$$\mathcal{H}_T = -\frac{\hbar^2}{2m}\nabla^2 + V_T(\mathbf{r}), \qquad r \in R_T \tag{3.2.2}$$

$$\mathcal{H}_S = -\frac{\hbar^2}{2m}\nabla^2 + V_S(\mathbf{r}), \qquad r \in R_S. \tag{3.2.3}$$

The difficulty arising from Eq. 3.2.1 is that an expression for the tunneling perturbation is not known exactly and moreover the unperturbed Hamiltonian H_0 doesn't seem to have any physical correspondence since it refers to two independent systems. In addition, even if an accurate description for a unified potential was available, it still double-counts the kinetic energy of the electrons. Therefore, a solution of the Schrodinger equation involving H_0 would not provide any physically relevant set of eigenfunctions and eigenvalues. Bardeen's suggestion was to introduce two sets of eigenfunctions $\{\chi_\nu\}$ and $\{\psi_\mu\}$ and eigenvalues $\{E_\nu\}$ and $\{E_\mu\}$ that would correspond to solutions of Schrodinger's equation for the tip and the surface independently. He then proceeded to describe the probability of a transition of an electron from a state μ in the tip to a state ν in the surface through Fermi's golden rule known from ordinary time-dependent perturbation theory

$$P = \frac{2\pi}{\hbar}\sum_\nu |M_{\mu\nu}|^2 \delta(E_\nu - E_\mu) \tag{3.2.4}$$

with the matrix element $M_{\mu\nu}$ given by

$$M_{\mu\nu} = \langle\psi_\mu|M|\chi_\nu\rangle = \int_{-\infty}^{+\infty} d\mathbf{r}\psi_\mu^*(H - H_S)\chi_\nu \tag{3.2.5}$$

and the Hamiltonian H

$$H = \begin{cases} H_T & \text{if } \mathbf{r} \in R_T \\ H_S & \text{if } \mathbf{r} \in R_S. \end{cases}$$

The physical justification of Eq. 3.2 lies on the fact that the tip and surface wavefunctions decay in the barrier region and are therefore nearly zero on the other side of the barrier. Hence, the Hamiltonian can be approximated by the independent tip and surface Hamiltonians in each case respectively. By exploiting the approximation 3.2, the tunneling matrix element $M_{\mu\nu}$ can be rewritten as

$$M_{\mu\nu} = \int_{R_S} d\mathbf{r}\psi_\mu^* \left(H_T - H_S\right) \chi_\nu \tag{3.2.6}$$

since in the tip region $H \sim H_T$ while in the surface region $H \sim H_S$ and therefore the integral 3.2.5 vanishes. It is possible to proceed by noting that 3.2.6 can be written as

$$M_{\mu\nu} = \int_{R_T} d\mathbf{r}\psi_\mu^* \left(H_T - H_S\right) \chi_\nu + \int_{R_T} d\mathbf{r}\chi_\nu (H_T - H_T)\psi_\mu^* \tag{3.2.7}$$

since the second term in 3.2.7 is simply zero. By re-arranging the terms, $M_{\mu\nu}$ becomes

$$M_{\mu\nu} = \int_{R_T} d\mathbf{r}\{\psi_\mu^* H_T \chi_\nu - \psi_\mu^* H_S \chi_\nu + \chi_\nu H_T \psi_\mu^* - \chi_\nu H_T \psi_\mu^*\}. \tag{3.2.8}$$

By virtue of Schrodinger's equation for the tip and the surface respectively

$$H_T\psi_\mu = E_\mu\psi_\mu$$

$$H_S\chi_\nu = E_\nu\chi_\mu$$

and therefore 3.2.8 becomes

$$M_{\mu\nu} = \int_{R_T} d\mathbf{r} \left\{\psi_\mu^* H_T \chi_\nu - \psi_\mu^* \chi_\nu E_\nu + \chi_\nu \psi_\mu^* E_\mu - \chi_\nu H_T \psi_\mu^*\right\}. \tag{3.2.9}$$

By neglecting all the other processes like, for example, inelastic scattering, and assuming that only elastic scattering takes place, $E_\mu = E_\nu$ and therefore

$$M_{\mu\nu} = \int_{R_T} d\mathbf{r} \left\{\psi_\mu^* H_T \chi_\nu - \chi_\nu H_T \psi_\mu^*\right\}. \tag{3.2.10}$$

Substituting the expression for the tip Hamiltonian to 3.2.10 $M_{\mu\nu}$ becomes

$$M_{\mu\nu} = \frac{\hbar^2}{2m} \int_{R_T} d\mathbf{r} \left\{\psi_\mu^* \nabla^2 \chi_\nu - \chi_\nu \nabla^2 \psi_\mu^* + \psi_\mu^* V_T(\mathbf{r}) \chi_\nu - \chi_\nu V_T(\mathbf{r}) \psi_\mu^*\right\} \tag{3.2.11}$$

and since the last two terms cancel

$$M_{\mu\nu} = \frac{\hbar^2}{2m} \int_{R_T} d\mathbf{r} \left\{ \psi_\mu^* \nabla^2 \chi_\nu - \chi_\nu \nabla^2 \psi_\mu^* \right\}. \tag{3.2.12}$$

Finally, by virtue of Green's theorem,

$$M_{\mu\nu} = \frac{\hbar^2}{2m} \int_S dS \left\{ \chi_\nu \nabla \psi_\mu^* - \psi_\mu^* \nabla \chi_\nu \right\}. \tag{3.2.13}$$

The expression for the tunneling current is eventually given in Bardeen's theory by

$$I = \frac{4\pi e}{\hbar} \sum_{\mu\nu} \left[f(E_\mu) - f(E_\nu + eV) \right] |M_{\mu\nu}|^2 \delta(E_\nu - E_\mu + eV) \tag{3.2.14}$$

with the tunneling matrix element $M_{\mu\nu}$ given by

$$M_{\mu\nu} = \frac{\hbar^2}{2m} \int_S dS \left\{ \chi_\nu \nabla \psi_\mu^* - \psi_\mu^* \nabla \chi_\nu \right\}. \tag{3.2.15}$$

ψ, χ are the electronic wavefunctions for the tip and the surface with energies E_ν and E_μ respectively while V is the bias voltage applied between the two. $f(E)$ is the Fermi distribution function. The integration surface S has to lie in the vacuum region between the tip and the surface and the summation extends over an energy interval around the Fermi energy. The advantage of Bardeen's method is that it can account for the electronic structure of both the tip and the surface and the way influences the tunneling current. In real-life applications, were STM images have to be simulated and compared with experiments for identification purposes, the wavefunctions for the tip and the surface can be obtained by electronic structure calculations using Density Functional Theory and utilizing Kohn-Sham orbitals.

Within the Tersoff-Hamann approximation [8, 9], the tip of the STM was treated within the spherical-tip approximation in which it is described by a single state ignoring the tip's internal structure. By employing Bardeen's formulation for the tunneling current and assuming that the tip is represented by a single s-state only, the tunneling current was proved to be proportional to the surface's density of states

$$I \sim \sum_\nu |\psi_\nu(\mathbf{r}_0)|^2 \delta(E_\nu - E_f) \tag{3.2.16}$$

were ψ_ν are the surface wavefunctions with energies E_ν and $\mathbf{r}_0$ is the position of the tip.

Although this is a computationally efficient approach which sometimes can elucidate experimental results at a qualitative level, it is eventually necessary to take into account how the tip's electronic structure affects the tunneling current and therefore the imaging of the surface. The Tersoff-Hamann approximation will be employed in a following chapter discussing the reactivity of the Rutile $TiO_2(110)$ surface and the implications on it due to the presence of charged oxygen vacancies.

Both the Bardeen and the Tersoff-Hamann approaches are now widely used in theoretical studies of surface properties, in conjunction with electronic structure calculations. The latter approach, in particular, is implemented in almost all major electronic structure codes. Both models have their merits with the Tersoff-Hamann approximation giving qualitative results at a low computational cost. The Bardeen approach, while treating both tip and surface on equal footing, is capable of providing results suitable for quantitative analysis and direct comparison with experimental data. Both approaches however do not include a complete treatment of the electronic transport problem, and they can only be considered as approximations in the low conductance, non-interacting regime. In the following section, an overview of the theoretical models behind the fully interacting electron transmission between the tip and the surface will be given. The interesting case of the coupling of the tunneling electrons to phonons in the interacting region will also be discussed in view of the recent advances in Inelastic Electron Tunneling Spectroscopy (IETS) experiments.

3.3 Scattering in the interaction picture

Consider a physical, many-body, system described by the time-independent Hamiltonian

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_i. \tag{3.3.1}$$

$\mathcal{H}_0$ is the free-particle part and H_i is the term incorporating all the various interactions between the particles. The free-particle component H_0 can be solved exactly to obtain the ground state $|\Phi_0(t)\rangle$ of the non-interacting system,

$$i\hbar\frac{d}{dt}|\Phi_0(t)\rangle = \mathcal{H}_0|\Phi_0(t)\rangle\,. \tag{3.3.2}$$

The system described by the Hamiltonian 3.3.1 is attached to a particle reservoir, and is assumed to be in thermal equilibrium. It can be described by the density matrix

$$\rho(\mathcal{H}) = \frac{e^{-\beta(\mathcal{H}-\mu N)}}{\mathrm{Tr}e^{-\beta(\mathcal{H}-\mu N)}}, \quad \beta = \frac{1}{kT}. \tag{3.3.3}$$

The energies are referenced to the chemical potential μ while $N = \sum_\nu c_\nu^\dagger c_\nu$ is the particle number operator. The dynamics of a quantum system can be described using three different frameworks, the Schrödinger picture, the Heisenberg picture and the interaction picture. They are equivalent between them and the decision regarding the adoption of the one over the other mainly depends to which representation exposes the physics in a more intuitive manner while simplifying the calculations at the same time. In the standard Schrödinger picture, the evolution of the system is driven by time-dependent wavefunctions while the operators remain constant,

$$\psi_S(\mathbf{r}, t) = U(t, t_0)\psi_S(\mathbf{r}, t_0), \quad \mathcal{O}_S = \text{const.} \tag{3.3.4}$$

with

$$U(t, t_0) = e^{-i\mathcal{H}(t-t_0)/\hbar} \tag{3.3.5}$$

the time evolution operator. In the case where the Hamiltonian has an explicit dependence on time, substituting Eq. 3.3.4 in the Schrödinger equation gives the general formal expression for the operator U,

$$U(t, t_0) \equiv 1 + \sum_{n=1}^{\infty} \left(-\frac{i}{\hbar}\right)^n \int_{t_0}^{t} dt_n \ldots \int_{t_0}^{t_3} dt_2 \int_{t_0}^{t_1} \mathcal{H}(t_n) \ldots \mathcal{H}(t_2)\mathcal{H}(t_1) = \tag{3.3.6}$$

$$= T\mathrm{exp}\left[-i \int_{t_0}^{t} dt' \mathcal{H}(t')\right] \tag{3.3.7}$$

which reduces to 3.3.5 for $\mathcal{H} = \text{const}$. In general,

$$U(t, t_0) = \begin{cases} T\mathrm{exp}\left[-i \int_{t_0}^{t} dt' \mathcal{H}(t')\right], & t \geq t_0 \\ \tilde{T}\mathrm{exp}\left[-i \int_{t_0}^{t} dt' \mathcal{H}(t')\right], & t < t_0 \end{cases}$$

where $T, \tilde{T}$ are the time and anti-time ordering operators that arrange the Hamiltonians in the exponentials into chronological order from right to left and left to right respectively. Substituting Eq. 3.3.4 in the time-dependent Schrödinger equation 3.3.2, the

kinetic equation for $U(t, t_0)$ is obtained,

$$i\hbar\frac{\partial U(t,t_0)}{\partial t} = \mathcal{H}(t)U(t,t_0). \tag{3.3.8}$$

Compared to the Schrödinger picture, the opposite approach holds true for the Heisenberg picture. In this representation, the time dependence of the wavefunctions is transferred to the operators,

$$\mathcal{O}_H(t) = e^{i\mathcal{H}t/\hbar}\mathcal{O}_S(0)e^{-i\mathcal{H}t/\hbar}, \quad \psi(\mathbf{r},t) = \text{const.} \tag{3.3.9}$$

The equivalent to Schrödinger's equation in the Heisenberg picture is

$$\frac{d}{dt}O_H(t) = \frac{i}{\hbar}\left[\mathcal{H}, \mathcal{O}_H(t)\right] + \left(\frac{\partial O_H}{\partial t}\right) \tag{3.3.10}$$

which is known as the Heisenberg equation of motion for the operator $\mathcal{O}_H$. The expectation values of physical observables are represented by Hermitian operators and do not depend on the selection of the picture. It therefore holds that

$$\langle\mathcal{O}\rangle = \langle\psi_S(t)|\mathcal{O}_S|\psi_S(t)\rangle = \langle\psi_H|U^\dagger(t,t_0)\mathcal{O}_S U(t,t_0)|\psi_H\rangle \tag{3.3.11}$$

$$= \langle\psi_H|\mathcal{O}_H(t)|\psi_H\rangle \tag{3.3.12}$$

with $|\psi_S(t)\rangle = U(t,t_0)\,|\psi_H\rangle$ relating the representations of the state vectors in the two pictures. The relation between the operators in the two pictures is given by

$$\mathcal{O}_H(t) = U^\dagger(t,t_0)\mathcal{O}_S U(t,t_0). \tag{3.3.13}$$

In framework, expectation values of observables represented by Hermitian operators O are given by

$$\langle O(t)\rangle = \text{Tr}\left[\rho O_{\mathcal{H}}(t)\right] \tag{3.3.14}$$

where ρ is the thermal equilibrium density matrix

$$\rho(H) = \frac{e^{-\beta H}}{\text{Tr}\left[e^{-\beta H}\right]}. \tag{3.3.15}$$

The middle ground between the two extremes of the Schrödinger and Heisenberg picture,

is the interaction picture. In the interaction picture, wavefunctions evolve under the interacting part of the Hamiltonian 3.3.1, $\mathcal{H}_i = \mathcal{H} - \mathcal{H}_0$, while the operators develop under the influence of the free-particle part H_0,

$$\psi_I(\mathbf{r}, t) = e^{-i\mathcal{H}_i t/\hbar}\psi_S(\mathbf{r}, 0) \tag{3.3.16}$$

$$\mathcal{O}_I(t) = e^{i\mathcal{H}_0 t/\hbar}\mathcal{O}_I(0)e^{-i\mathcal{H}_0 t/\hbar}. \tag{3.3.17}$$

The relation between the wavefunctions in the interaction picture and in the Schrödinger picture can be found by replacing in Eq. 3.3.16 the interaction part $\mathcal{H}_i$ of the Hamiltonian with $\mathcal{H} - \mathcal{H}_\prime$. The factor $e^{-i\mathcal{H}t/\hbar}$ multiplying $\psi_S(\mathbf{r}, 0)$ develops the wavefunction in time in the Schrödinger picture, $\psi_S(\mathbf{r}, t) = e^{-i\mathcal{H}t/\hbar}\psi_S(\mathbf{r}, 0)$ while $e^{i\mathcal{H}_\prime t/\hbar}$ alters the representation from $\psi_S(\mathbf{r}, t)$ to $\psi_I(\mathbf{r}, t)$. For an arbitrary initial time t_0, the relation between the usual Schrödinger picture for the wavefunction and the interaction picture representation becomes,

$$\psi_I(\mathbf{r}, t) = e^{i\mathcal{H}_0(t-t_0)/\hbar}\psi_S(\mathbf{r}, t_0). \tag{3.3.18}$$

By defining the operator $U_0(t, t_0) = e^{-i\mathcal{H}_0(t-t_0)/\hbar}$, Eq. 3.3.18 can be re-casted in the form $\psi_I(\mathbf{r}, t) = U_0^\dagger(t, t_0)\psi_S(\mathbf{r}, t_0)$ with $U_0^\dagger = U_0^{-1}$ and $U_0(t_0, t_0) = 1$. The operator $U_0(t, t_0)$ satisfies the following equation of motion,

$$i\hbar\frac{\partial}{\partial t}U_0 = \mathcal{H}_0 U_0 \tag{3.3.19}$$

with $\mathcal{H}_0$ in the Schrödinger picture. By starting from the time-dependent Schrödinger equation and substituting $\psi_S(\mathbf{r}, t)$ with 3.3.18, it can also be shown that wavefunctions in the interaction picture obey

$$i\hbar\frac{\partial}{\partial t}\psi_I(\mathbf{r}, t) = V_I(t)\psi_I(\mathbf{r}, t). \tag{3.3.20}$$

Following the same argument on the invariance of expectation values of physical operators regarding their pictorial representation, the interaction picture equivalent of 3.3.13 is

$$\mathcal{O}_I(t) = U_0^\dagger(t, t_0)\mathcal{O}_S U_0(t, t_0). \tag{3.3.21}$$

It should be observed that in the interaction picture, both wavefunctions and operators

have an explicit dependence on time. For the representation to be complete, an equation of motion for the operators is also needed in addition to Eq. 3.3.20. This equation can be proved to be

$$i\hbar\frac{\partial}{\partial t}\mathcal{O}_i(t) = [\mathcal{O}_I(t), \mathcal{H}_0] + i\hbar U_0^\dagger(t,t_0)\frac{\partial}{\partial t}\mathcal{O}_S U_0(t,t_0). \tag{3.3.22}$$

Returning to the representation of wavefunctions in the interaction picture, the operator $U(t)$ defined by $\psi_I(t) = U(t)\psi_S(0)$ is

$$U(t) = e^{i\mathcal{H}_0 t}e^{-i\mathcal{H}t/\hbar}. \tag{3.3.23}$$

The wavefunction $\psi_I(t)$ at a later time $t > t'$ combined with $\psi_I(t') = U(t')\psi_S(0)$, equals

$$\psi_I(t) = U(t)U^\dagger(t')\psi_I(t') = \mathcal{S}(t,t')\psi_I(t'). \tag{3.3.24}$$

The S-matrix $\mathcal{S}(t,t') = \mathcal{S}^\dagger(t',t) = U(t)U^\dagger(t')$ describes the evolution of the wavefunc-

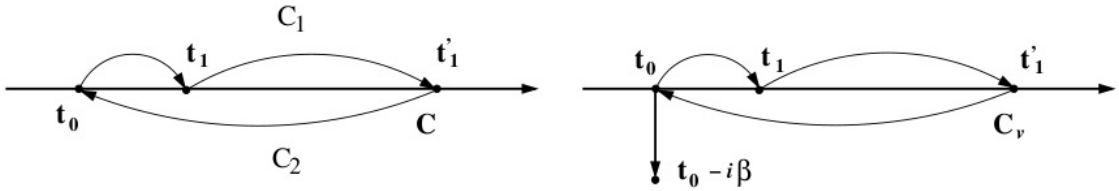

Figure 3.2: Time contour in the perturbative expansion.

tion, in the interaction picture, between two arbitrary moments t and t', with t at a later point compared to t' on the time axis. By substituting $\mathcal{S}(t,t')$ in Eq. 3.3.20 and performing the integration, the S-matrix can be obtained to be

$$S(t,t') = T\exp\left[-\frac{i}{\hbar}\int_{t'}^{t} dt_1 V_I(t_1)\right]. \tag{3.3.25}$$

This expression can be expanded in a MacLaurin series to give

$$S(t,t') = 1 + \sum_{n=1}^{\infty}\frac{1}{n!}\left(\frac{-i}{\hbar}\right)^n \int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2 \ldots \int_{t_0}^{t_n - 1} dt_n T\left[V_I(t_1)V_I(t_2)\ldots V_I(t_n)\right]. \tag{3.3.26}$$

Substituting Eq. 3.3.26 back in Eq. 3.3.24, a perturbative expansion of the wavefunction at time t is obtained. Eq. 3.3.26 represents in a perturbative manner multiple scattering during the propagation of the particle by the potential centre $V_I(t)$ [10]. One can change

between the Heisenberg and interaction picture representations for the operator O by noting that

$$O_H(t) = e^{i\mathcal{H}t}e^{-i\mathcal{H}_0 t}O_I(t)e^{i\mathcal{H}_0 t}e^{-i\mathcal{H}t} \tag{3.3.27}$$

or, in terms of the S-matrix operator,

$$O_H(t) = S(0,t)O_I(t)S(t,0). \tag{3.3.28}$$

This discussion regarding the Heisenberg and the interaction picture along with the following presentation of the many-body Green's functions, provides the fundamental building blocks in presenting non-equilibrium theory and the formalism behind the evaluation of the tunneling current in out-of-equilibrium junctions.

3.4 Many-Body Green's Functions

Consider the single-particle, time-independent Hamiltonian

$$\mathcal{H} = \mathcal{H}_0 + V \tag{3.4.1}$$

in which $\mathcal{H}_0$ is the free-particle Hamiltonian and V a perturbation. The solution to the time-dependent Schrödinger's equation

$$i\hbar\frac{\partial\psi(t)}{\partial t} = \mathcal{H}\psi(t) \tag{3.4.2}$$

or, alternatively,

$$\left[i\frac{\partial}{\partial t} - \mathcal{H}_0 - V\right]\psi(\mathbf{r},t) = 0 \tag{3.4.3}$$

can proceed using the Green's functions technique [11] familiar from the eigenvalue problem of linear algebra. The Green's functions are defined by

$$\left[i\hbar\frac{\partial}{\partial t} - \mathcal{H}_0\right]\mathcal{G}_0(\mathbf{r}t,\mathbf{r}'t) = \delta(\mathbf{r}-\mathbf{r'})\delta(t-t') \tag{3.4.4}$$

$$\left[i\hbar\frac{\partial}{\partial t} - \mathcal{H}_0 - V(\mathbf{r})\right]\mathcal{G}(\mathbf{r}t,\mathbf{r}'t) = \delta(\mathbf{r}-\mathbf{r'})\delta(t-t') \tag{3.4.5}$$

$$\tag{3.4.6}$$

and therefore,

$$\mathcal{G}_0(\mathbf{r}t,\mathbf{r}'t) = \left[i\hbar\frac{\partial}{\partial t} - \mathcal{H}_0\right]^{-1} \tag{3.4.7}$$

$$\mathcal{G}(\mathbf{r}t,\mathbf{r}'t) = \left[i\hbar\frac{\partial}{\partial t} - \mathcal{H}_0 - V(\mathbf{r})\right]^{-1}. \tag{3.4.8}$$

The physical meaning of the Green's function is that it has the property of propagating the wavefunction in space-time [10]. For a known wavefunction $\psi(\mathbf{r}_0,t_0)$ at $(\mathbf{r}_0,t_0)$, the wavefunction at a later point in space-time $(\mathbf{r},t)$ is given by

$$\psi(\mathbf{r},t) = \int d\mathbf{r}_0 \mathcal{G}(\mathbf{r}t,\mathbf{r}_0t_0)\psi(\mathbf{r}_0,t_0). \tag{3.4.9}$$

with probability $|\mathcal{G}(\mathbf{r}t,\mathbf{r}_0t_0)|^2$. It can be seen, by direct substitution, that the following two solutions satisfy Eq. 3.3.16,

$$\psi(\mathbf{r},t) = \psi_0(\mathbf{r},t) + \int d\mathbf{r}' \int dt' \mathcal{G}_0(\mathbf{r}t,\mathbf{r}'t')V(\mathbf{r}')\psi(\mathbf{r}',t') \tag{3.4.10}$$

$$\psi(\mathbf{r},t) = \psi_0(\mathbf{r},t) + \int d\mathbf{r}' \int dt' \mathcal{G}(\mathbf{r}t,\mathbf{r}'t')V(\mathbf{r}')\psi_0(\mathbf{r}',t') \tag{3.4.11}$$

$$\tag{3.4.12}$$

and by recursion lead to Dyson's equation,

$$\mathcal{G} = \mathcal{G}_0 + \mathcal{G}_0 V \mathcal{G}. \tag{3.4.13}$$

In Eq. 3.4.13, the spatial and time integrals have been suppressed to simplify the notation. By observing Eq. 3.4.9, it is easy to see that the propagation of the particle depends on its initial position $\mathbf{r}_0$ at time t_0. No explicit assumption is made on whether time t is before or after time t_0 on the time axis. It is therefore possible to introduce two Green's functions depending on the relative position of t with respect to t'. In this case t' replaces t_0 in the notation to abstract the arrangement of time in the equations. The retarded Green's function $\mathcal{G}^r(\mathbf{r}t,\mathbf{r}'t')$ propagates the particle from an earlier time t' to a later time t while the advanced Green's function $\mathcal{G}^a$ does the opposite i.e $t' > t$. The retarded and advanced Green's functions are given in real-space by

$$\mathcal{G}^r(\mathbf{r}t,\mathbf{r}'t') = -i\theta(t-t')\left\langle\mathbf{r}|e^{-i\mathcal{H}(t-t')}|\mathbf{r}'\right\rangle \tag{3.4.14}$$

$$\mathcal{G}^a(\mathbf{r}t,\mathbf{r}'t') = i\theta(t'-t)\left\langle\mathbf{r}|e^{-i\mathcal{H}(t-t')}|\mathbf{r}'\right\rangle. \tag{3.4.15}$$

The discussion of Green's functions now switches to the many-body case. Although conceptually the similarities with the single-particle formalism discussed in the previous paragraph are many, the difference lies in the fact that now Green's functions although defined for single particles, are governed by the full many-body Hamiltonian. Within the many-body framework, the Green's functions for a free particle system represented by the Hamiltonian $\mathcal{H} = \sum_{\mathbf{k}\sigma} \epsilon_{\mathbf{k}\sigma} c^{\dagger}_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma}$ are defined as follows:

$$\mathcal{G}_0^{>}(\mathbf{k}\sigma, t-t') = -i(1-n_F(\epsilon_{\mathbf{k}}))e^{-\epsilon_{\mathbf{k}}(t-t')} \tag{3.4.16}$$

$$\mathcal{G}_0^{<}(\mathbf{k}\sigma, t-t') = in_F(\epsilon_{\mathbf{k}})e^{-\epsilon_{\mathbf{k}}(t-t')} \tag{3.4.17}$$

$$\mathcal{G}_0^{r}(\mathbf{k}\sigma, t-t') = -i\theta(t-t')e^{-\epsilon_{\mathbf{k}}(t-t')} \tag{3.4.18}$$

$$\mathcal{G}_0^{a}(\mathbf{k}\sigma, t-t') = i\theta(t'-t)e^{-\epsilon_{\mathbf{k}}(t-t')}. \tag{3.4.19}$$

These definitions can be extended to the interacting case and are summarised below [11–13],

$$\mathcal{G}^{>}(\mathbf{r}\sigma t, \mathbf{r}'\sigma' t') = -i\left\langle \psi_{\sigma}(\mathbf{r}t)\psi^{\dagger}_{\sigma}(\mathbf{r}'t')\right\rangle \tag{3.4.20}$$

$$\mathcal{G}^{<}(\mathbf{r}\sigma t, \mathbf{r}'\sigma' t') = -i(\pm 1)\left\langle \psi_{\sigma}(\mathbf{r}t)\psi^{\dagger}_{\sigma}(\mathbf{r}'t')\right\rangle \tag{3.4.21}$$

$$\mathcal{G}^{r}(\mathbf{r}\sigma t, \mathbf{r}'\sigma' t') = -i\theta(t-t')\left\langle \left[\psi_{\sigma}(\mathbf{r},t), \psi^{\dagger}_{\sigma'}(\mathbf{r}'t')\right]_{B,F}\right\rangle \tag{3.4.22}$$

$$\mathcal{G}^{a}(\mathbf{r}\sigma t, \mathbf{r}'\sigma' t') = i\theta(t'-t)\left\langle \left[\psi_{\sigma}(\mathbf{r},t), \psi^{\dagger}_{\sigma'}(\mathbf{r}'t')\right]_{B,F}\right\rangle \tag{3.4.23}$$

where the brackets denote the usual commutation (anti-commutation) relations for bosons (fermions) . One can also define

$$\mathcal{G}(\mathbf{r},t;\mathbf{r}',t') = -\frac{i}{\hbar}\left\langle T_t\left[\psi(\mathbf{r},t)\psi^{\dagger}(\mathbf{r},t)\right]\right\rangle = \theta(t-t')\mathcal{G}^{>}(\mathbf{r}t,\mathbf{r}'t') + \theta(t'-t)\mathcal{G}^{<}(\mathbf{r}t,\mathbf{r}'t') \tag{3.4.24}$$

and

$$\tilde{\mathcal{G}}(\mathbf{r},t;\mathbf{r}',t') = -\frac{i}{\hbar}\left\langle \tilde{T}_t\left[\psi(\mathbf{r},t)\psi^{\dagger}(\mathbf{r},t)\right]\right\rangle = \theta(t-t')\mathcal{G}^{<}(\mathbf{r}t,\mathbf{r}'t') + \theta(t'-t)\mathcal{G}^{>}(\mathbf{r}t,\mathbf{r}'t') \tag{3.4.25}$$

where the operators T and $\tilde{T}$ are the time and anti-time ordering operators.

An example of the formalism presented in this section is the quantum mechanical tunneling of electrons in a metal-insulator-metal junction. The STM is an example of such a junction with the role of the insulator to be played by the vacuum region between

the tip and the surface. In its simplest form, such a system consists of two contacts in equilibrium with electrons flowing, due to the application of a bias voltage, from the one contact to the other without interacting either between themselves or with the environment. In such a case, the Hamiltonian is given by

$$\begin{aligned}\mathcal{H} &= \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{H}_T \\ &= \sum_{\mu} \epsilon_{1,\mu} c^{\dagger}_{1,\mu} c_{1,\mu} + \sum_{\mu} \epsilon_{2,\mu} c^{\dagger}_{2,\mu} c_{2,\mu} + \sum_{\mu\nu} \left(T_{\mu\nu} c^{\dagger}_{1,\mu} c_{2,\nu} + T^{*}_{\mu\nu} c^{\dagger}_{2,\nu} c_{1,\mu} \right), \end{aligned} \tag{3.4.26}$$

with

$$T_{\mu\nu} = \int d\mathbf{r} \psi^{*}_{\mu}(\mathbf{r}) \mathcal{H}(\mathbf{r}) \psi_{\nu}(\mathbf{r}) \tag{3.4.27}$$

the tunneling matrix describing the coupling between the two contacts. The tunneling current is the rate of change of electrons occupying states in the leads i.e

$$I_e = -e \langle \frac{dN_1}{dt} \rangle = -ie \langle [\mathcal{H}_T, N_1] \rangle \tag{3.4.28}$$

with $N_1 = \sum_{\nu} c^{\dagger}_{\nu} c_{\nu}$ the particle number operator. The tunneling current between the two contacts can be derived to be [11]

$$I = \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi} \sum_{\mu\nu} |T_{\mu\nu}|^2 \left[G^{>}_1(\nu, \epsilon) G^{<}_2(\mu, \epsilon + eV) - G^{<}_1(\nu, \omega) G^{>}_2(\mu, \omega + eV) \right] \tag{3.4.29}$$

with V the bias voltage i.e the potential difference between the two leads. This is an oversimplified model of the tunneling of electrons from the tip to the surface in the Scanning Tunneling Microscope and it was discussed only as a brief example of the power of the Green's functions within the formalism of non-interacting systems. The proper treatment of the STM will be discussed in the sections below using the full formalism of non-equilibrium theory. It will lead to an exact calculation of Eq. 3.4.28 incorporating both elastic and inelastic scattering in the barrier region. Before proceeding however, the key concepts behind the non-equilibrium theory will be discussed to provide the correct background theory for such an evaluation.

3.5 Non-equilibrium Theory

The fully interacting, time-dependent Hamiltonian 3.3.1, unlike the non-interacting counterpart $\mathcal{H}_0$ cannot be solved directly to obtain the ground state $|\Psi_0\rangle$ of the system.

The approach due to Gell-Mann and Low [14] was to adopt an adiabatic approximation in which the interacting part of the Hamiltonian is slowly switched on and then back off, taking the system from a non-interacting distant past, to a fully interacting system at $t = t_0$, and then back to a non-interacting distant future

$$\mathcal{H} = \mathcal{H}_0 + e^{-\epsilon|t|}\mathcal{V}(0). \tag{3.5.1}$$

The benefit of this approach was that the ground state $|\Psi_0\rangle$ of the fully-interacting system can now be obtain by evolving the non-interacting ground state $|\Phi_0\rangle$ in time, from the infinite past to the interacting period t_0 and then to the infinite future

$$|\Psi_0(t_0)\rangle = S(t_0, -\infty)\,|\Phi_0\rangle\,, \quad |\Phi_0\rangle = S(\infty, t_0)\,|\Psi_0\rangle\,. \tag{3.5.2}$$

The S-matrix describing this transition is given by

$$S(\infty, -\infty) = S(\infty, t_0)S(t_0, -\infty) = T\text{exp}\left[-i\int_{-\infty}^{\infty} dt_1 V(t_1)\right] \tag{3.5.3}$$

which by expansion to a Taylor series is re-casted to the form

$$S(\infty, -\infty) = \sum_{n=0}^{\infty} \frac{(-i)^{n+1}}{n!} \int_{-\infty}^{\infty} dt_1 \ldots dt_n T\left\{V(t_1)\ldots V(t_n)\right\}. \tag{3.5.4}$$

The Green's function describing the propagation of an electron under the influence of the Hamiltonian $\mathcal{H}$ is given in this framework by

$$\mathcal{G}(x,t;x',t') = -i\frac{\langle\Phi_0|T\left\{S(\infty,-\infty)\psi(x,t)\psi^\dagger(x',t')\right\}|\Phi_0\rangle}{\langle\Phi_0|S(\infty,-\infty)|\Phi_0\rangle} \tag{3.5.5}$$

or by using 3.5.4 in 3.5.5 and in virtue of Wick's theorem [15],

$$\mathcal{G}(x,t;x',t') = -i\sum_{n=0}^{\infty}(-i)^n \int_{-\infty}^{+\infty} dt_1 \ldots dt_n \,\langle\Phi_0|\psi(x,t)\psi^\dagger(x',t')V(t_1)\ldots V(t_n)|\Phi_0\rangle_{\text{conn}} \tag{3.5.6}$$

involving only all the topologically nonequivalent connected Feynman diagrams [16].

When the system is driven by the interaction out of equilibrium, it cannot be assumed that the adiabatic approach will take the system from the ground state $|\Phi_0\rangle$ at $t \to -\infty$ to the same ground state $|\Phi_0\rangle$ as $t \to \infty$. The non-equilibrium processes break the symmetry between distant past and distant future, rendering the formalism developed

above non-applicable directly. To surpass this problem, it was suggested that the S-matrix 3.3.25 should evaluated on a time contour with two branches, the first branch C_1 evolving from $t = -\infty$ to t_0 and the other branch, C_2, from t_0 back to $t = -\infty$ as in Fig. 3.3. The system therefore is evolved adiabatically from a well known non-interacting ground state $|\Phi_0\rangle$ to the fully interacting $|\Psi_0\rangle$ and then back to $|\Phi_0\rangle$, despite the broken symmetry between $t = -\infty$ and $t = \infty$ [17]. Under this ordering, the S-matrix 3.3.25 can be decomposed to two different components, each one evaluated on a different branch on the contour

$$S(t, t_0) = T_{C_1} \exp\left[-\frac{i}{\hbar}\int_{t_0}^{t} dt' V(t')\right] \tag{3.5.7}$$

$$S(t_0, t) = T_{C_2} \exp\left[-\frac{i}{\hbar}\int_{t}^{t_0} dt' V(t')\right] \tag{3.5.8}$$

The introduction of the Schwinger contour now allows for expectation values of observables to be evaluated in the out-of-equilibrium case. Considering the time evolution of the density matrix operator 3.3.15 in the interaction picture,

$$\rho_I(t) = \mathcal{S}_I(t, t_0)\rho(t_0)\mathcal{S}_I^{\dagger}(t, t_0) \tag{3.5.9}$$

with the S-matrix in the interaction picture given as in 3.3.25 by

$$\mathcal{S}_I(t, t_0) = T \exp\left[-i\int_{t_0}^{t} dt V_I(t)\right] \tag{3.5.10}$$

$$\mathcal{S}_I^{\dagger}(t, t_0) = T^{\dagger} \exp\left[+i\int_{t_0}^{t} dt V_I(t)\right], \tag{3.5.11}$$

it is now possible to introduce a time ordering operator T_c which orders time labels along the two branches of the contour. The density matrix operator can then be re-casted in the form

$$\rho_I(t) = T_C \exp\left[-i\int_{C} dt V_I(t)\right] \rho(t_0) \tag{3.5.12}$$

and evaluated on the contour $C = C_1 \cup C_2$ instead. Expectation values of operators can then be defined on the contour as

$$\langle O \rangle = \left\langle T_C \exp\left[-i\int_{C} d\tau V_I(\tau)\right] O_I(t)\right\rangle_0 \tag{3.5.13}$$

where $\langle \ldots \rangle_0$ is the thermal average with respect to $\rho(t_0) = \exp(-\beta\mathcal{H}) / \mathrm{Tr}[\exp(-\beta\mathcal{H})]$. It can be seen that $\rho(t_0)$ includes the fully interacting, time independent Hamiltonian

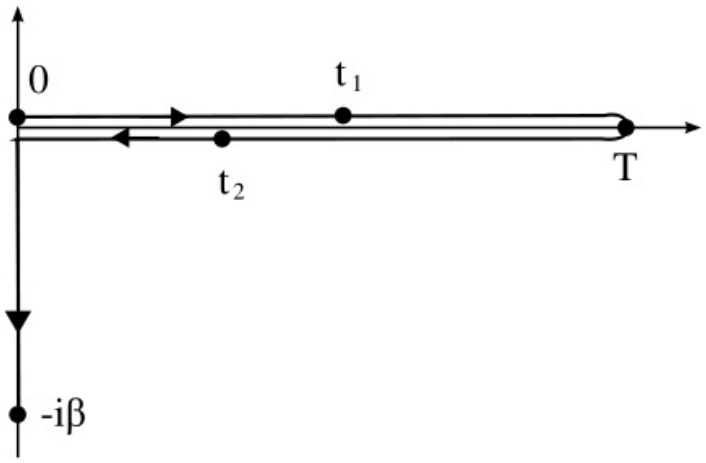

Figure 3.3: The Schwinger time contour. The upper branch runs from $t = -\infty$ to a time T while the lower branch from T back to $-\infty$.

$\mathcal{H}$. By utilising a second interaction picture in which the time dependence is transferred to $\mathcal{H}$

$$\mathcal{H}(\tau) = e^{i\mathcal{H}_0\tau}\mathcal{H}e^{-i\mathcal{H}_0\tau}, \tag{3.5.14}$$

the expectation value 3.5.13 can be written with respect to the ground state charge density ρ_0

$$\left\langle T_C \exp\left\{-i\int_C d\tau\left[\mathcal{H}(\tau)+V(\tau)\right]\right\}O(t)\right\rangle_0. \tag{3.5.15}$$

Time τ in Eq. 3.5.15 runs on the Schwinger contour C and the thermal average in this case is with respect to the charge density of the non-interacting system at $t = -\infty$ represented by the Hamiltonian $\mathcal{H}_0$. It is now clear that non-equilibrium theory, initially difficult to treat because of the asymmetry between the ground state $|\Phi_0\rangle$ of the non-interacting system and the ground state $|\Psi_0\rangle$ of the fully interacting system, has now been mapped to the well understood equilibrium case. This was achieved by introducing a time contour C running from $(t = -\infty, \mathcal{H}_0)$ to $(t_0, \mathcal{H})$, and then back to the original starting point, $(t = -\infty, \mathcal{H}_0)$. By introducing a new time ordering operator T_C which orders time labels on the contour, all the quantities of interest can now be evaluated as in the standard equilibrium case by substituting real-time integrals with integrals along the branches of the contour C. In this case, the contour-ordered Green's function is given by

$$\mathrm{G}(t_1, t_2) \equiv -\frac{i}{\hbar}\left\langle T_C\left[\psi(t_1)\psi^\dagger(t_2)\right]\right\rangle \tag{3.5.16}$$

satisfying the Dyson equation

$$\mathcal{G}(x_1,t_1;x_\prime,t_\prime) = \mathcal{G}_0(t_1,t_2) + \int d^3x_2 \int_C d\tau_2 \mathcal{G}(x_1,t_1;x_2,t_2)U(x_2,t_2)\mathcal{G}(x_2,t_2;x^{'},t^{'}) \tag{3.5.17}$$

$$+ \int d^3x_2 \int d^3x_3 \int_c d\tau_2 \int_c d\tau_3 \mathcal{G}_0(x_1,t_1;x_2,t_2)\Sigma(x_2,t_2;x_3,t_3)\mathcal{G}(x_3,t_3;x^{'},t^{'}) \tag{3.5.18}$$

with the non-equilibrium term in the Hamiltonian assuming the form

$$H(t) = \int dx \psi^\dagger(x)U(x,t)\psi(x). \tag{3.5.19}$$

3.6 The Keldysh Formulation

If initial correlations are not important and can be ignored, a further simplification can be made by allowing each branch of the time contour to span the entire time axis. This is achieved by letting $t_0 \to \infty$ and is equivalent to dropping the t_0 to $t_0 - i\beta$ branch of the contour in Fig. 3.3 [18]. The new contour now has two branches, C_1 and C_2, running from $-\infty$ to ∞ and from ∞ to $-\infty$ respectively. This formalism, due to Keldysh

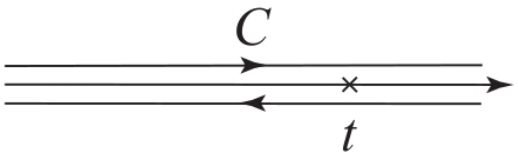

Figure 3.4: The Keldysh time contour. The upper branch runs from $-\infty$ to ∞ while the lower branch from ∞ to ∞.

[19], introduces a different time-ordering operator, which orders time labels according to their position on the two different branches of the Keldysh time contour. The Keldysh contour ordered Green's function now branches out to four different functions:

$$\mathcal{G}(t,t') = \begin{cases} \mathcal{G}^{>}(t_1,t_2), & t_1 \in C_2,\, t_2 \in C_1 \\ \mathcal{G}^{<}(t_1,t_2), & t_1 \in C_1,\, t_2 \in C_2 \\ \mathcal{G}^{T}(t_1,t_2), & t_1,t_2 \in C_1 \\ \mathcal{G}^{\tilde{T}}(t_1,t_2), & t_1,t_2 \in C_2. \end{cases}$$

$\mathcal{G}^{>}$ and $\mathcal{G}^{<}$ are the greater and lesser Green's functions while $\mathcal{G}^{T}$ and $\mathcal{G}^{\tilde{T}}$ the time-ordered and anti-time-ordered Green's functions of the Keldysh formalism. They are given by

$$\mathcal{G}^{T}(t_1,t_2) = -i\left\langle T\hat{\psi}(t_1)\hat{\psi}^{\dagger}(t_2)\right\rangle \tag{3.6.1}$$

$$= -i\theta(t_1-t_2)\left\langle \hat{\psi}(t_1)\hat{\psi}^{\dagger}(t_2)\right\rangle + \theta(t_2-t_1)\left\langle \hat{\psi}^{\dagger}(t_2)\hat{\psi}(t_1)\right\rangle \tag{3.6.2}$$

$$\mathcal{G}^{\tilde{T}}(t_1,t_2) = -i\left\langle \tilde{T}\hat{\psi}(t_1)\hat{\psi}^{\dagger}(t_2)\right\rangle \tag{3.6.3}$$

$$= -i\theta(t_2-t_1)\left\langle \hat{\psi}(t_1)\hat{\psi}^{\dagger}(t_2)\right\rangle + \theta(t_1-t_2)\left\langle \hat{\psi}^{\dagger}(t_2)\hat{\psi}(t_1)\right\rangle \tag{3.6.4}$$

$$\mathcal{G}^{>}(t_1,t_2) = -i\left\langle \hat{\psi}(t_1)\hat{\psi}^{\dagger}(t_2)\right\rangle \tag{3.6.5}$$

$$\mathcal{G}^{<}(t_1,t_2) = i\left\langle \hat{\psi}^{\dagger}(t_2)\hat{\psi}(t_1)\right\rangle . \tag{3.6.6}$$

It can be seen that $\mathcal{G}^{T}+\mathcal{G}^{\tilde{T}} = \mathcal{G}^{>}+\mathcal{G}^{<}$ giving three linearly independent functions and therefore allowing the freedom to define the more suitable advanced

$$\mathcal{G}^{a}(t_1,t_2) = i\theta(t_2-t_1)\left\langle\{\psi(t_1),\psi(t_2)\}\right\rangle \tag{3.6.7}$$

$$= \theta(t_2-t_1)\left[\mathcal{G}^{<}(t_1,t_2)-\mathcal{G}^{>}(t_1,t_2)\right] \tag{3.6.8}$$

and retarded

$$\mathcal{G}^{r}(t_1,t_2) = -i\theta(t_2-t_1)\left\langle\{\psi(t_1),\psi(t_2)\}\right\rangle \tag{3.6.9}$$

$$= \theta(t_2-t_1)\left[\mathcal{G}^{>}(t_1,t_2)-\mathcal{G}^{<}(t_1,t_2)\right] \tag{3.6.10}$$

Green's functions, for which $\mathcal{G}^{r}-\mathcal{G}^{a} = \mathcal{G}^{>}-\mathcal{G}^{<}$.

By observing Dyson's equation 3.5.17, it is apparent that products of operators need to be evaluated on the Keldysh contour C. Several rules, due to Langreth [20], exist for evaluating integrals of the forms appearing in Dyson's equation, by employing analytical continuation for projections to the real time axis. Considering the integral below,

$$C(\tau_1,\tau_2) = \int_C d\tau A(\tau_1,\tau)B(\tau,\tau_3) \tag{3.6.11}$$

defined on the Keldysh contour C, it can be evaluated by deforming the contour C to two different contours C_1 and C_2 as in Fig. 3.2. Using $C^{<}$ as an example, 3.6.11

becomes

$$C^{<}(t_1, t_{1'}) = \int_{C_1} d\tau A(t_1, \tau) B^{<}(\tau, t_{1'}) + \int_{C_{1'}} d\tau A^{<}(t_1, \tau) B(\tau, t_{1'}). \tag{3.6.12}$$

By choosing $t_1 \in C_1$ then the integration over the contour C_1 in 3.6.12 can be divided

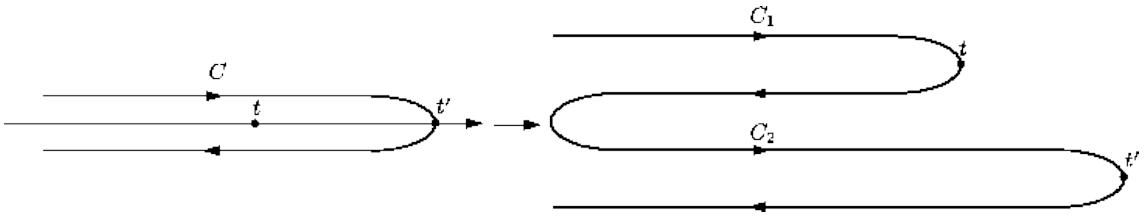

Figure 3.5: Deformation of the time contour.

in an integration over the region $[-\infty, t_1]$ and the region $[t_1, \infty]$ as follows

$$\int_{C_1} A(t_1, \tau) B^{<}(\tau, t_{1'}) = \int_{-\infty}^{t_1} dt A^{>}(t_1, t) B^{<}(t, t_{1'}) + \int_{t_1}^{-\infty} dt A^{<}(t_1, t) B^{<}(t, t_{1'}) \tag{3.6.13}$$

$$= \int_{-\infty}^{\infty} dt A^{r}(t_1, t) B^{<}(t, t_{1'}) \tag{3.6.14}$$

in virtue of $A \equiv A^{>}(t_1, t)$ in $[-\infty, t_1]$ and $A \equiv A^{<}(t_1, t)$ in $[t_1, \infty]$. The second term in 3.6.12 is evaluated, following a similar line of thought, to be

$$\int_{C_{1'}} d\tau A^{<}(t_1, \tau) B(\tau, t_{1'}) = \int_{-\infty}^{\infty} A^{<}(t_1, t) B^{a}(t, t_{1'}). \tag{3.6.15}$$

By putting 3.6.13 and 3.6.15 in 3.6.12 and by applying the same technique for $C^{>}$ as with $C^{<}$, the following expressions for the greater and lesser functions can be obtained,

$$C^{>}(t_1, t_2) = \int_{-\infty}^{+\infty} dt \left[A^{r}(t_1, t) B^{>}(t, t_2) + A^{>}(t_1, t) B^{a}(t, t_2) \right] \tag{3.6.16}$$

$$C^{<}(t_1, t_2) = \int_{-\infty}^{+\infty} dt \left[A^{r}(t_1, t) B^{<}(t, t_2) + A^{<}(t_1, t) B^{a}(t, t_2) \right]. \tag{3.6.17}$$

For the advanced and retarded Green's functions,

$$C^{a}(t_1, t_2) = \int_{-\infty}^{+\infty} dt A^{a}(t_1, t) B^{a}(t, t_2) \tag{3.6.18}$$

$$C^{r}(t_1, t_2) = \int_{-\infty}^{+\infty} dt A^{r}(t_1, t) B^{r}(t, t_2) \tag{3.6.19}$$

When integrals on the contour consist of products of three functions as below

$$D(\tau_1,\tau_2)=\int_C d\tau d\tau^{'}A(\tau_1,\tau)B(\tau,\tau_{\prime})C(\tau^{'},\tau) \tag{3.6.20}$$

the analytic continuation leads to the following expressions

$$D^{<}=\int dt\left[A^rB^rC^{<}+A^rB^{<}C^a+A^{<}B^aC^a\right] \tag{3.6.21}$$

$$D^{>}=\int dt\left[A^rB^rC^{>}+A^rB^{>}C^a+A^{>}B^aC^a\right]. \tag{3.6.22}$$

Products of functions with both time variables running on the contour C,

$$C(\tau,\tau')=A(\tau,\tau')B(\tau,\tau') \tag{3.6.23}$$

$$C(\tau,\tau')=A(\tau,\tau')B(\tau',\tau) \tag{3.6.24}$$

acquire the following form on the real-time axis [21]

$$C^{<,>}(t_1,t_2)=A^{<,>}(t_1,t_2)B^{<,>}(t_1,t_2) \tag{3.6.25}$$

$$D^{<,>}(t_1,t_2)=A^{<,>}(t_1,t_2)B^{>,<}(t_2,t_1 \tag{3.6.26}$$

while the advanced and retarded counterparts become

$$C^r(t_1,t_2)=A^{<}(t_1,t_2)B^r(t_1,t_2)+A^r(t_1,t_2)B^{<}(t_1,t_2)+A^r(t_1,t_2)B^r(t_1,t_2) \tag{3.6.27}$$

$$D^r(t_1,t_2)=A^r(t_1,t_2)B^{<}(t_2,t_1)+A^{<}(t_1,t_2)B^a(t_2,t_1) \tag{3.6.28}$$

$$=A^{<}(t_1,t_2)B^a(t_2,t_1)+A^r(t_1,t_2)B^{<}(t_2,t_1) \tag{3.6.29}$$

The rules 3.6.16, 3.6.20, and 3.6.21 can be summarised by the Keldysh matrices

$$\hat{A}=\begin{pmatrix} A^r & A^{<,>} \\ 0 & A^a \end{pmatrix}$$

which allow them to be re-casted in the form

$$\hat{D}=\hat{A}\hat{B} \tag{3.6.30}$$

and

$$\hat{D}=\hat{A}\hat{B}\hat{C}. \tag{3.6.31}$$

It is now possible to return to the Dyson equation 3.5.17, with the contour C on which the integrals are to be evaluated to the Keldysh contour 3.4. Langreth's rules can then be applied to give the Keldysh equations,

$$\mathcal{G}^{<} = (1 + \mathcal{G}^{r}\Sigma^{r})\,\mathcal{G}_{0}^{<}\,(1 + \Sigma^{a}\mathcal{G}^{a}) + \mathcal{G}^{r}\Sigma^{<}\mathcal{G}^{a} \tag{3.6.32}$$

$$\mathcal{G}^{r,a} = \mathcal{G}_{0}^{r,a}\Sigma^{r,a}\mathcal{G}^{r,a}. \tag{3.6.33}$$

To summarise this section, it can be noted that the fundamental result has been that theory behind non-equilibrium Green's functions has the same structure as the equilibrium theory at $t = 0$. This was achieved by passing from time-ordered Green's functions defined on a real-time axis, to contour ordered Green's functions. The contour ordering operator T_c orders time labels on the different branches of the contour. The original proposal for the contour C involved branches running from $-\infty$ to T and back. A further simplification was, however, achieved by neglecting initial correlation phenomena taking place at $t = -\infty$, and by therefore allowing the branches to span the entire axis i.e from $t = -\infty$ to $t = \infty$ as in Fig. 3.4. Dyson's equation can then be written on the contour and after specifying a self-energy Σ, Langreth's rules for analytic continuation can be employed to define the advanced/retarded and greater/lesser self-energy components. Those can then be used in evaluating the coupled Keldysh equations for the calculation of the equivalent components of the non-equilibrium Green's function. The machinery presented in this section has provided the theoretical framework for the study of electron transport in interacting, out-of-equilibrium, mesoscopic systems. The calculation of the tunneling current in those systems and the relation to the STM will be discussed in the next section.

3.7 Electron Transport in Interacting Mesoscopic Systems

The Scanning Tunneling Microscope belongs to a larger class of systems that can be modeled as two infinite metal leads, acting as electron reservoirs, and an interacting region in between. In the case of the STM, the interacting region is taken to be the vacuum region between the tip and the surface. The vacuum region can also include

molecules or cluster of atoms [4]. Such systems are described by the Hamiltonian

$$\mathcal{H} = \mathcal{H}_C + \mathcal{H}_T + \mathcal{H}_D, \tag{3.7.1}$$

where $\mathcal{H}_C = \mathcal{H}_L + \mathcal{H}_R$ is the Hamiltonian representing the left and the right lead, $\mathcal{H}_T$ represents the coupling between the leads and the interacting, or device, region, while $\mathcal{H}_D$ is the Hamiltonian corresponding to the interactions in the region between the two leads. At $t = -\infty$ the two leads are in equilibrium with chemical potentials μ_L and μ_R for the left and the right lead respectively. A shift in the chemical potential caused by the application of a bias voltage takes the system out of equilibrium. Electrical current starts flowing between the two leads and after an initial period in which correlations are dominant, the system reaches a steady state. In second quantization, Eq. 3.7.1 can be written using single-particle annihilation and creation operators for the electrons in the individual regions. It is,

$$\mathcal{H} = \sum_{k,a\epsilon L,R} \epsilon_{ka} c^\dagger_{ka} c_{ka} + \sum_{k,a\epsilon L,Rn} \left[V_{ka,n} c^\dagger_{ka} d_n + \text{h.c.} \right] + \mathcal{H}_\text{D} \left(\{ d^\dagger_n \}; \{ d_n \} \right) \tag{3.7.2}$$

This Hamiltonian is the starting point for calculating the tunneling current flowing between the two leads. The operators $\{c^\dagger\}, \{c\}$ are the creation and annihilation operators for the electrons in the leads while $\{d^\dagger\}, \{d\}$ similarly for the interacting region. $\mathcal{H}_C = \sum_{k,a\epsilon L,R} \epsilon_{ka} c^\dagger_{ka} c_{ka}$ is the Hamiltonian representing the uncoupled leads, with Green's functions given by

$$g^<_{k\alpha}(t-t') = i \langle c^\dagger_{k\alpha}(t') c_{k\alpha}(t) \rangle = i f(\epsilon^0_{k\alpha}) \text{exp} \left[-i\epsilon^0_{k\alpha}(t-t') \right] \tag{3.7.3}$$

$$g^{r,a}_{k\alpha}(t-t') = \mp i\theta(\pm t \mp t') \langle \{ c_{k\alpha}(t), c^\dagger_{k\alpha}(t') \} \rangle = \mp i\theta(\pm t \mp t') \text{exp} \left[-i\epsilon_{k\alpha}(t-t') \right]. \tag{3.7.4}$$

In this section g instead of G_0 will be used to denote the free-particle Green's functions in order to simplify the notation. The component $\mathcal{H}_T$ of the full Hamiltonian describes the tunneling of the electrons from the leads to the central region. The form of $\mathcal{H}_\text{D}$ is not specified explicitly as it can vary depending on the nature of the interactions in the device. If the electrons do not interact then it is simply $\mathcal{H}_\text{D} = \sum_m \epsilon_m d^\dagger_m d_m$. If there is coupling to phonons then, for a single electronic state, it is

$$\mathcal{H}_\text{D} = \epsilon_0 d^\dagger d + d^\dagger d \sum_q M_q \left[a^\dagger_q + a_{-q} \right] \tag{3.7.5}$$

with $a^\dagger, a$ the creation and annihilation operators for phonons in the central region with $\sum_q \hbar\omega_q a_q^\dagger a_q$ the additional free-phonon contribution to the total Hamiltonian $\mathcal{H}$.

The tunneling current operator for the left lead L is given by

$$I_L = -e\,\langle \frac{d\mathcal{N}_L}{dt}\rangle = -\frac{ie}{\hbar}\,\langle[\mathcal{H},\mathcal{N}_L]\rangle \tag{3.7.6}$$

$$= -\frac{ie}{\hbar}\sum_{k,n}\Big[-V_{kL,n}c_{kL}^\dagger d_n + V_{kL,n}^* d_n^\dagger c_{kL}\Big] \tag{3.7.7}$$

with $\mathcal{N}_L = \sum_{k,a\epsilon L} c_{ka}^\dagger c_{ka}$ the number of electrons propagating through the contact L to the central region. In virtue of Kirchoff's law, the total tunneling current flowing is then given by

$$\mathcal{J} = \frac{1}{2}\left(\mathcal{J}^L - \mathcal{J}^R\right) \tag{3.7.8}$$

with a similar to Eq. 3.7.2 expression holding for $\mathcal{J}^R$. The calculation proceeds by evaluating $\mathcal{J}^L$ and $\mathcal{J}^R$ using 3.7.2 and 3.7.6 and then by substituting them in Eq. 3.7.8. The relevant to the discussion Green's functions where originally obtained by Caroli et al. [22] in their study for the tunneling current in metal-insulator-metal junctions. The Keldysh lesser Green's functions was derived to be

$$\mathcal{G}_{n,k\alpha}^<(t,t_1) = i\,\langle c_{k\alpha}^\dagger(t) d_n(t_1)\rangle \tag{3.7.9}$$

$$\mathcal{G}_{k\alpha,n}^<(t,t_1) = i\,\langle d_n^\dagger(t_1) c_{k\alpha}(t)\rangle\,. \tag{3.7.10}$$

The Green's functions $G_{n,k\alpha}^< = -\left[G_{k\alpha,n}^<\right]^*$ are the single-particle propagators for the electrons tunneling from the lead $\alpha \in L, R$ to an intermediate state n in the central region. They obey the Dyson equations

$$\mathcal{G}_{ka,n}^<(\omega) = \sum_m V_{ka,m}\left[g_{ka,ka}^t(\omega)\mathcal{G}_{m,n}^<(\omega) - g_{ka,ka}^<(\omega)\mathcal{G}_{m,n}^{\tilde{t}}(\omega)\right] \tag{3.7.11}$$

$$\mathcal{G}_{n,ka}^<(\omega) = \sum_m V_{ka,m}^*\left[g_{ka,ka}^<(\omega)\mathcal{G}_{n,m}^t(\omega) - g_{ka,ka}^{\tilde{t}}(\omega)\mathcal{G}_{n,m}^<(\omega)\right], \tag{3.7.12}$$

where $t, \tilde{t}$ denote Keldysh time ordered and anti-time ordered operators. Similarly with 3.7.9, the following Green's functions describing electron propagation from a state n to

another, m, in the central region can also be defined,

$$\mathcal{G}^{<}_{n,m}(t,t_1) = i\left\langle d^{\dagger}_m(t) d_n(t_1)\right\rangle \tag{3.7.13}$$

$$\mathcal{G}^{>}_{n,m}(t,t_1) = -i\left\langle d_n(t_1) d^{\dagger}_m(t)\right\rangle . \tag{3.7.14}$$

The tunneling current from the left lead is now evaluated to be,

$$\langle I_L\rangle = \mathcal{J}_L = \frac{ie}{\hbar}\sum_{k,a\epsilon L}\left(V_{ka,n}\left\langle c^{\dagger}_{ka} d_n\right\rangle - V^{*}_{ka,n}\left\langle d^{\dagger}_n c_{ka}\right\rangle\right) \tag{3.7.15}$$

$$= \frac{e}{\hbar}\sum_{k,\alpha\in L}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\left[V_{k\alpha,n}G^{<}_{n,k\alpha}(\omega) - V^{*}_{k\alpha,n}G^{<}_{k\alpha,n}(\omega)\right]. \tag{3.7.16}$$

Following Meir and Wingreen [23], the tunneling current for the left lead can be obtained by substituting 3.7.13 in 3.7.11 to obtain $\mathcal{G}^{<}_{k\alpha,n}$ by using $\mathcal{G}^{>}(\omega)+\mathcal{G}^{<}(\omega) = \mathcal{G}^{t}(\omega)+\mathcal{G}^{\tilde{t}}(\omega)$, $\mathcal{G}^{>}(\omega) - \mathcal{G}^{<}(\omega) = \mathcal{G}^{r}(\omega) - \mathcal{G}^{a}(\omega)$ and the unperturbed Green's functions

$$g^{<}_{ka,ka}(\omega) = 2\pi i f_L(\omega)\delta(\omega-\epsilon_{ka}) \tag{3.7.17}$$

$$g^{>}_{ka,ka}(\omega) = -2\pi i\left[1-f_L(\omega)\right]\delta(\omega-\epsilon_{ka}). \tag{3.7.18}$$

After substituting 3.7.11 in 3.7.15, the tunneling current is given by

$$\mathcal{J}_L = \frac{ie}{\hbar}\sum_{a\epsilon L,n,m}\int d\epsilon n_a(\epsilon)V_{a,n}(\epsilon)V^{*}_{a,m}(\epsilon)\left\{f_L(\epsilon)\left[G^{R}_{n,m}(\epsilon) - G^{A}_{n,m}(\epsilon)\right] + G^{<}_{n,m}(\epsilon)\right\}. \tag{3.7.19}$$

Using a similar expression with 3.7.19 for the right contact and by substituting in Eq. 3.7.8, the total tunneling current flowing through the interacting region is given by [23]

$$J = \frac{ie}{2\hbar}\int\frac{d\epsilon}{2\pi}\mathrm{Tr}\left\{\left[f_L(\epsilon)\Gamma^{L}(\epsilon) - f_R(\epsilon)\Gamma^{R}(\epsilon)\right]\left(G^{R}(\epsilon) - G^{A}(\epsilon)\right) + \left(\Gamma^{L}(\epsilon) - \Gamma^{R}(\epsilon)\right)\mathcal{G}^{<}(\epsilon)\right\} \tag{3.7.20}$$

with Γ^{L}, Γ^{R} the contacts, representing the coupling of the leads to the interacting region. They are defined by

$$\Gamma^{L}_{nm} = 2\pi\sum_{a\epsilon L} n_a(\epsilon)V_{an}(\epsilon)V^{*}_{am}(\epsilon) \tag{3.7.21}$$

$$\Gamma^{R}_{nm} = 2\pi\sum_{a\epsilon R} n_a(\epsilon)V_{an}(\epsilon)V^{*}_{am}(\epsilon) \tag{3.7.22}$$

$$\tag{3.7.23}$$

The Landauer's formula, [11, 24, 25], for the tunneling current in the non-interacting regime can be derived from Eq. 3.7.20 using the following relations for the Green's functions

$$\mathcal{G}^{<} = if_L(\epsilon)\mathcal{G}^R\Gamma^L\mathcal{G}^A + if_R(\epsilon)\mathcal{G}^R\Gamma^R\mathcal{G}^a \tag{3.7.24}$$

$$\mathcal{G}^r - \mathcal{G}^a = -i\mathcal{G}^r(\Gamma^L + \Gamma^R)\mathcal{G}^a. \tag{3.7.25}$$

By substituting 3.7.24 and 3.7.25 to Eq. 3.7.20 the tunneling current becomes

$$J = \frac{e}{2\pi\hbar}\int \frac{d\epsilon}{2\pi}\left[f_L(\epsilon) - f_R(\epsilon)\right]\mathrm{Tr}\left\{G^a\Gamma^R G^r\Gamma^L\right\} \tag{3.7.26}$$

or

$$J = \frac{e}{2\pi\hbar}\int \frac{d\epsilon}{2\pi}\left[f_L(\epsilon) - f_R(\epsilon)\right]\mathrm{Tr}\left[tt^{\dagger}\right] \tag{3.7.27}$$

with t being the transmission coefficient between channels a and $a^{'}$ of the two leads. The transmission coefficient is defined by

$$t_{a,a'} = 2\pi\sum_{n,m} n_a^{1/2} n_{a'}^{1/2} V^*_{a,n} G^r_{n,m} V_{a',m}. \tag{3.7.28}$$

It is now possible to compare the Landauer formula given in Eq. 3.7.26 with the tunneling current for the fully interacting system 3.7.20. By using the definition of the advanced and retarded self-energies

$$\Sigma^r = \left[(g^r)^{-1} - G^r\right]^{-1} \tag{3.7.29}$$

$$\Sigma^a = \left[(g^a)^{-1} - G^a\right]^{-1} \tag{3.7.30}$$

along with

$$G^r - G^a = G^r\Sigma G^a \tag{3.7.31}$$

$$\Sigma = \Sigma^r - \Sigma^a. \tag{3.7.32}$$

Eq. 3.7.20 can be re-casted in the form

$$J = \frac{e}{2\pi\hbar}\int \frac{d\epsilon}{2\pi}\left[f_L(\epsilon) - f_R(\epsilon)\right]\mathrm{Tr}\left\{G^a\Gamma^R G^r\Gamma^L\Sigma_0^{-1}\Sigma\right\}. \tag{3.7.33}$$

It can be seen that the term $\Sigma_0^{-1}\Sigma$ appearing in the Meir-Wingreen formula for the

tunneling current provides the qualitative difference between Eq. 3.7.20 and the Landauer formula for the current. In the interacting case, Σ includes all those contributions including electron-electron and electron-phonon interactions that shift Σ from its equilibrium value, i.e $\Sigma \neq \Sigma_0$. When, however, only elastic scattering is present, $\Sigma = \Sigma_0$ and Eq. 3.7.20 reduces to the Landauer formula.

The expression for the tunneling current given by Eq. 3.7.20 can be generalised to accommodate the time-dependent case for which the current is explicitly dependent on time [26]. In contrast to the steady-state case in which the expression of the current can be antisymmetrised in virtue of Kirchoff's law, the explicit dependence on time does not, in this case, allow for a similar manipulation. The expression for the current flowing from the left lead is,

$$\mathcal{J}_L(t) = -\frac{2e}{\hbar}\int_{-\infty}^{t} dt_1 \int \frac{d\epsilon}{2\pi} \mathrm{Im}\mathrm{Tr}\left\{e^{-i\epsilon(t_1-t)}\Gamma^L(\epsilon,t_1,t)\left[G^<(t,t_1) + f_L(\epsilon)G^r(t,t_1)\right]\right\} \tag{3.7.34}$$

It can be seen that by taking the time-independent limit $\Gamma(\epsilon,t_1,t) \to \Gamma(\epsilon)$ and performing the integrations, Eq. 3.7.19 and 3.7.20 are retrieved. In the case where the couplings of the contacts to the central region are proportional to each other, the expression for the current assumes the form [23, 26]

$$\mathcal{J} = \frac{ie}{\hbar}\int \frac{d\epsilon}{2\pi}\left[f_L(\epsilon) - f_R(\epsilon)\right]\mathrm{Tr}\left\{\frac{\Gamma^L(\epsilon)\Gamma^R(\epsilon)}{\Gamma^L(\epsilon)+\Gamma^R(\epsilon)}\left[\mathcal{G}^R(\epsilon) - \mathcal{G}^a(\epsilon)\right]\right\}. \tag{3.7.35}$$

The formalism presented in the last two sections allowed for the derivation of an expression for the tunneling current 3.7.20, 3.7.34 in out-of-equilibrium conditions when interactions between electrons are present. They are valid in the case where the leads are non-interacting, and are formally exact. They reduce to the standard Bardeen's equation for the tunneling current in the case where no interactions are present and therefore the system is described by the simple Hamiltonian $\mathcal{H} = \sum_{k,\alpha\epsilon L,R} \epsilon_{k,a} c^\dagger_{k,\alpha} c_{k,\alpha}$ [4]. Inelastic scattering effects introduced by the presence of molecules in the contact region are also accounted for, due to the renormalisation they introduce in the self-energy. It is the presence of the self-energy and the difficulty in defining an expression that captures all the essential physics that poses the challenge in making the theory practical. In the current state-of-the-art codes, the dominant approach is to use a real-space representation in which the non-equilibrium Green's functions, the contacts and the self-energies

are evaluated using iterative schemes involving Kohn-Sham orbitals within the Density Functional Theory framework [4].

3.8 Tunneling Current within Plane-Wave DFT

In the previous section, the exact formalism using non-equilibrium theory and Green's functions was briefly reviewed. The exact Meir-Wingreen formula [23, 26] for the tunneling current in a mesoscopic system was presented. The derivation of this formula was based on early theoretical work done by Schwinger [17], Kadanoff and Baym [12], Abrikosov et al. [27], Langreth [20], and Keldysh [19] among others, as well as Caroli et al. [22] who later on applied the formalism to electron transport and inelastic scattering in metal-insulator-metal junctions. In this section, contact will be made with Density Functional Theory and the mechanism for evaluating tunneling currents within a plane-wave framework will be presented. This will unify the discussion regarding Density Functional Theory presented in the previous chapter with the STM and will offer the basic framework which was used for performing the STM simulations in the study of the reactivity of Rutile $TiO_2(110)$ and the structure of the $TiO_2(011)$ surfaces.

In a STM experiment, the tunneling current is of the order of 1 nA and the time interval between single-electron processes of 10^{-10} sec. It is therefore possible to argue that since the probability of finding two electrons in the barrier simultaneously is very low, electron-electron interactions in the vacuum region between the leads can safely be neglected [4]. Also, lattice excitations do not happen within these timescales and can therefore be neglected. The tunneling current in this case is given by the Landauer-Büttiker formula 3.7.26, presented again below,

$$J = \frac{e}{2\pi\hbar} \int \frac{d\epsilon}{2\pi} \left[f_L(\epsilon) - f_R(\epsilon) \right] \mathrm{Tr} \left\{ \mathcal{G}^a \Gamma^R \mathcal{G}^r \Gamma^L \right\}. \tag{3.8.1}$$

The STM is considered to be represented by two infinite metal leads representing the tip and the surface, and a vacuum barrier in between. The tip and the surface have chemical potentials μ_S and μ_T respectively. The application of the bias voltage V has the effect of shifting the chemical potentials of the tip and the surface by $e\mathrm{V}/2$. The system is initially in equilibrium, with $\mu_S = \mu_T = \mu_0$. The application of the bias voltage drives the system out of equilibrium, the chemical potentials of the tip and the

surface are shifted, $\mu_S = \mu_0 - eV/2$ and $\mu_T = \mu_0 + eV/2$, and current begins to flow from the tip to the surface. The Landauer formula Eq. 3.8.1 can then be re-written as

$$J = \frac{e}{2\pi\hbar}\int \frac{d\epsilon}{2\pi}\left[f_S(\mu_S,\epsilon) - f_T(\mu_T,\epsilon)\right]\mathrm{Tr}\left\{\mathcal{G}^a\Gamma_R\mathcal{G}^r\Gamma_S\right\} \tag{3.8.2}$$

where $f(\mu,\epsilon)$ is the Fermi distribution, $\mathcal{G}^{r(a)}(\epsilon)$ the retarded (advanced) Green's function of the barrier region and Γ_T, Γ_S the tip and surface contacts respectively, representing the coupling of the tip and the surface to the barrier. There exist methods for evaluating the tunneling current within the above framework using localised atomic orbitals as basis functions [28]. The prevalence of plave-wave DFT codes[1] and the limited accuracy of the atomic orbitals caused by a rapid decay of the wavefunctions in the barrier region make it advantageous to recast the formulation within a plane-wave formalism.

By treating the tip and the surface within plane-wave DFT independently, the Kohn-Sham orbitals can be obtained and used to evaluate the advanced and retarded Green's functions for the tip and the surface. The Green's functions can then be written as [4],

$$\mathcal{G}_S^{r(a)}(\mathbf{r}_1,\mathbf{r}_2,\epsilon) = \sum_i \frac{\psi_i(\mathbf{r}_1)\psi_i^*(\mathbf{r}_1)}{\epsilon - \epsilon_i^{'} + (-)i\eta} \tag{3.8.3}$$

$$\mathcal{G}_T^{r(a)}(\mathbf{r}_1,\mathbf{r}_2,\epsilon) = \sum_j \frac{\chi_j(\mathbf{r}_1)\chi_j^*(\mathbf{r}_1)}{\epsilon - \epsilon_j^{'} + (-)i\eta'} \tag{3.8.4}$$

where $\{\psi\}, \{\chi\}$ are the Kohn-Sham orbitals for the surface and tip respectively. It should be noted here that as the Kohn-Sham states do not have a clear physical meaning, using them to construct the advanced and retarded Green's functions for the tip and the surface is clearly an approximation, necessitated by the lack of better alternatives for $\{\psi\}$ and $\{\chi\}$. The energy eigenvalues $\epsilon_i^{'} = \epsilon_i - eV/2$ and $\epsilon_j^{'} = \epsilon_j + eV/2$ are shifted under the presence of the bias voltage. The contacts are defined as

$$\Gamma_S = 2\eta\sum_k \psi_k(\mathbf{r}_3)\psi_k^*(\mathbf{r}_4), \quad \Gamma_T = 2\eta'\sum_i \chi_k(\mathbf{r}_1)\chi_k^*(\mathbf{r}_2). \tag{3.8.5}$$

In this framework one proceeds by arguing that the retarded and advanced Green's functions in the barrier region are given to the first order by [4]

$$\mathcal{G}_0^{r(a)}(\mathbf{r}_1,\mathbf{r}_2,\epsilon) = \mathcal{G}_S^{r(a)}(\mathbf{r}_1,\mathbf{r}_2,\epsilon) + \mathcal{G}_T^{r(a)}(\mathbf{r}_1,\mathbf{r}_2,\epsilon). \tag{3.8.6}$$

[1] Within the physicists' community at least.

The Landauer formula can then be evaluated by substituting Eq. 3.8.6 and 3.8.5 in 3.8.2. The trace after taking the limit $\eta, \eta' \to +0$ becomes

$$\mathrm{Tr}\left\{\mathcal{G}_0^a \Gamma_R \mathcal{G}_0^r \Gamma_S\right\} = 4\eta\eta' \sum_{ik} |A_{ik}|^2 4\pi^2 \delta(\epsilon - \epsilon_k')\delta(\epsilon - \epsilon_i')(\epsilon - \epsilon_k' + \epsilon - \epsilon_i')^2 \tag{3.8.7}$$

with A_{ik} the overlap integral

$$A_{ik} = \int d^3 r \chi^*(\mathbf{r})\psi_k(\mathbf{r}) = \frac{1}{k_i^2 - k_k^2} \int dS \left[\chi_i^*(\mathbf{r})\nabla\psi_k(\mathbf{r}) - \psi_k(\mathbf{r})\nabla\chi_i^*(\mathbf{r})\right]. \tag{3.8.8}$$

Substituting Eq. 3.8.8 in 3.8.7 and by using

$$\epsilon_i = \frac{\hbar^2 k_i^2}{2m} = \epsilon_i' - \frac{eV}{2}, \quad \epsilon_k = \frac{\hbar^2 k_k^2}{2m} = \epsilon_k' - \frac{eV}{2} \tag{3.8.9}$$

for the energy eigenvalues, the tunneling current to zero-order becomes

$$\begin{aligned} \mathcal{J}_0 = \frac{4\pi e}{\hbar} \sum_{ik} \Big[& f(\mu_S, \epsilon_k - \frac{eV}{2}) - f(\mu_T, \epsilon_i + \frac{eV}{2}) \Big] \\ & \times \left| \left(-\frac{\hbar^2}{2m} - \frac{eV}{k_i^2 - k_k^2} \right) M_{ik} \right|^2 \delta(\epsilon_i - \epsilon_k + eV). \end{aligned} \tag{3.8.10}$$

M_{ik} is the matrix element

$$M_{ik} = \int dS \left[\chi_i^*(\mathbf{r})\nabla\psi_k(\mathbf{r}) - \psi_k(\mathbf{r})\nabla\chi_i^*(\mathbf{r})\right]. \tag{3.8.11}$$

It can be seen from Eq. 3.8.10 that the bias voltage induces a correction to the tunneling current given by Bardeen's formula Eq. 3.2.14. Indeed, by setting $V = 0$, Bardeen's equation for the tunneling current is retrieved confirming the fact that the full scattering treatment reduces to Bardeen's formula in the zero bias limit [29, 30].

The formulation presented above can be extended to any order in virtue of Dyson's equation. To first-order,

$$\mathcal{G}_1^{r(a)} = \mathcal{G}_0^{r(a)} + \mathcal{G}_0^{r(a)} V \mathcal{G}_0^{r(a)} \tag{3.8.12}$$

giving for weakly-coupled limit $V = V_T + V_S$

$$\mathcal{G}_1^{r(a)} = \mathcal{G}_0^{r(a)} - \frac{\hbar^2}{m} \sum_{i,k} \frac{\psi_i(\mathbf{r}_1) M_{ki}^* \chi_k^*(\mathbf{r}_2) + \chi_i(\mathbf{r}_1) M_{ik} \psi_k^*(\mathbf{r}_2)}{(\epsilon - \epsilon_i' \pm i\eta)(\epsilon - \epsilon_k' \pm i\eta')}. \tag{3.8.13}$$

Similarly to Eq. 3.8.12 and 3.8.13 the Green's functions can be evaluated to any order,

usually keeping a balance between accuracy and computational cost [4].

3.9 Coupling to Phonons - IETS

For completeness, it should be mentioned at this point that the framework that unifies DFT with STM simulations as presented in the previous section, has also been extended to incorporate electron-phonon interactions in the barrier region. The field of inelastic electron tunneling spectroscopy has grown in the last decade, since it was initiated by studying the vibrational spectra of acetylene on Cu(100) [31, 32]. The task, however, of interpreting the results still remains a difficult one due to the complexity of the technique. It has been therefore a necessity to develop computational methods for the calculation of the tunneling current when electrons couple to vibrational modes in the barrier region. Electron-phonon interactions in the electron transport problem had been studied before the recent advances in the experimental techniques [22], and expressions for the tunneling current adapted to accommodate them have been derived [22, 33]. More particularly, however, recent developments have allowed for the evaluation of the tunneling current within a ground-state Density Functional Theory framework, in conjunction with the many-body Bardeen's theory for the tunneling current [34, 35]. The tunneling current is, within this framework, given by

$$\mathcal{J}(V) = \frac{2e^2}{\hbar\pi}\left(\frac{\hbar^2}{2m}\right)^2 \int_{\epsilon_F}^{\epsilon_F+eV} d\omega \mathrm{Tr}\left[\left(\overleftarrow{\nabla}_1 - \overrightarrow{\nabla}_1\right)\mathrm{Im}\mathcal{G}_T^r(\mathbf{r}_1,\mathbf{r}_2,\omega)\right] \times\left[\left(\overleftarrow{\nabla}_2 - \overrightarrow{\nabla}_2\right)\mathcal{G}_S^r(\mathbf{r}_2,\mathbf{r}_1,\omega)\right] \tag{3.9.1}$$

where $\mathcal{G}_{T(S)}^{r(a)}$ are the retarded (advanced) Green's functions for the tip and the surface as given by Eq. 3.8.3. Local vibrations are introduced as perturbations,

$$\mathcal{H}_{ph} = \sum_{\mu\nu} u'_{\mu\nu} c_\mu^\dagger c_\nu \delta Q(b^\dagger - b) \tag{3.9.2}$$

where $\delta Q = \sqrt{\hbar/(2M\Omega)}$ with Ω the vibration frequency of the localised mode. The Keldysh greater Green's function $\mathcal{G}^>$ for the surface can be written using a one-electron basis set as follows

$$\mathcal{G}_S^>(\mathbf{r},\mathbf{r}',\omega) = -2i\pi\sum_\lambda (1-f_\lambda)\psi_\lambda(\mathbf{r})\psi_\lambda^*(\mathbf{r}')\delta(\omega-\epsilon_k). \tag{3.9.3}$$

with $f_\lambda = n_F(\epsilon_\lambda)$ the Fermi distribution and $\psi_\lambda(\mathbf{r})$ the single-particle Kohn-Sham orbitals for the surface. Within perturbation theory, 3.9.3 acquires the first-order correction $\delta\mathcal{G}^>_S = \delta\mathcal{G}^>_{\text{inel}} + \delta\mathcal{G}^>_{\text{el}}$ with

$$\delta\mathcal{G}^>_{\text{inel}}(\mathbf{r},\mathbf{r}',\omega) = \int d^3r_1 d^3r_2 \mathcal{G}^r(\mathbf{r},\mathbf{r}_1,\omega)\Sigma^<(\mathbf{r}_1,\mathbf{r}_2,\omega)\mathcal{G}^a(\mathbf{r}_2,\mathbf{r}',\omega) \tag{3.9.4}$$

$$\begin{aligned}\delta\mathcal{G}^>_{\text{el}}(\mathbf{r},\mathbf{r}',\omega) = &\int d^3r_1 d^3r_2 \left[\mathcal{G}^>(\mathbf{r},\mathbf{r}_1,\omega)\Sigma^a(\mathbf{r}_1,\mathbf{r}_2,\omega)\mathcal{G}^a(\mathbf{r}_2,\mathbf{r}',\omega)\right] \\ &+ \int d^3r_1 d^3r_2 \left[\mathcal{G}^r(\mathbf{r},\mathbf{r}_1,\omega)\Sigma^r(\mathbf{r}_1,\mathbf{r}_2,\omega)\mathbf{G}^>(\mathbf{r}_2,\mathbf{r}',\omega)\right].\end{aligned} \tag{3.9.5}$$

with $\delta\mathcal{G}^>_{\text{inel}}, \delta\mathcal{G}^>_{\text{el}}$ being the inelastic and elastic contributions respectively to the correlation function $\mathcal{G}^>$ respectively. The missing ingredients for evaluating these contributions are the self-energies, they are given by

$$\Sigma^>(\mathbf{r}_1,\mathbf{r}_2,\omega) = -2i\pi u(\mathbf{r}_1)u(\mathbf{r}_2)\sigma_\lambda(1-f_\lambda)\psi_\lambda(\mathbf{r}_1)\psi^*_\lambda(\mathbf{r}_2)\delta(\omega-\Omega-\epsilon_\lambda) \tag{3.9.6}$$

for the greater self-energy $\Sigma^>$. For the advanced self-energy, the expression is

$$\Sigma^a(\mathbf{r}_1,\mathbf{r}_2,\omega) = u(\mathbf{r}_1)u(\mathbf{r}_2)\psi_\lambda(\mathbf{r}_1)\psi^*_\lambda(\mathbf{r}_2)\times\left[\frac{1-f_\lambda}{\omega-\Omega-\epsilon_\lambda-i\delta}+\frac{f_\lambda}{\omega+\Omega-\epsilon_\lambda-i\delta}\right] \tag{3.9.7}$$

with $\Sigma^r = [\Sigma^a]^*$. The delta function in 3.9.6 - and the lack of it in 3.9.7 - represents the fact that the exchange of the quantum of energy during the scattering appears, as expected by definition, in the inelastic correction to $\mathcal{G}^>$ only.

By using $\mathcal{G}^>(\mathbf{r}_1,\mathbf{r}_2,\omega) = 2i(1-n_F(\omega))\text{Im}\mathcal{G}^r(\mathbf{r}_1,\mathbf{r}_2,\omega)$, substituting 3.9.3-3.9.7 and 3.8.3 in 3.9.1 and differentiating with respect to V, the change in conductance induced by the coupling to phonon modes can be extracted. The inelastic contribution is given by

$$\begin{aligned}\delta\left(\frac{dI}{dV}(\omega)\right)_{\text{ine}} =& \frac{\pi}{2}\sum_{m,\lambda}(1-f_\lambda)\times\left|\int d\mathbf{S}\left(\delta\psi_\lambda\overrightarrow{\nabla}\psi^*_m - \psi_m\overrightarrow{\nabla}\delta\psi^*_\lambda\right)\right|^2 \\ &\times\delta(\epsilon_m-\omega)\delta(\omega-\Omega-\epsilon_\lambda)\end{aligned} \tag{3.9.8}$$

with the perturbed wavefunction given by

$$\delta\psi_\lambda = \sum_\mu \frac{\langle\mu|u|\lambda\rangle}{\epsilon_\lambda-\epsilon_\mu+i\delta}. \tag{3.9.9}$$

It can be seen that Eq. 3.9.8 is a Bardeen-like equation satisfied by the perturbed sample wavefunctions 3.9.10. In virtue of the delta functions in 3.9.8, there is a sharp

peak at $eV = \hbar\Omega$ and $T = 0$. This is the fundamental concept behind IETS which is captured within this framework. The bias voltage V can be varied, and for $V = V_0$, when it satisfies $eV_0 = \hbar\Omega$, a peak appears in the dI/dV spectrum therefore allowing the identification of Ω. The elastic contribution in the conductance is more difficult to treat, however it is a sufficient approximation to use

$$\delta\left(\frac{dI}{dV}(\omega)\right)_{\text{el}} = -\delta\left(\frac{dI}{dV}(\omega)\right)_{\text{ine}} \tag{3.9.10}$$

with the perturbed sample wavefunctions now given by

$$\delta\psi_\lambda = \sqrt{2}\pi \sum_\mu \psi_\mu \langle\mu|u|\lambda\rangle \, \delta(\epsilon_\lambda - \epsilon_\mu). \tag{3.9.11}$$

STM simulations within the frameworks presented in this and the previous sections, are now systematically used in conjunction with DFT to quantitatively compare and supplement experimental results of adsorbates on surfaces [35, 36].

3.10 Conclusions

In this chapter, the theoretical framework for evaluating the tunneling current in a Scanning Tunneling Microscope setup was briefly reviewed. Expressions for the tunneling current where presented within different approximations, from the simple Tersoff-Hamann approach [8, 9] to the full treatment using non-equilibrium theory and Green's functions [22, 23]. The Scanning Tunneling Microscope is today extensively used for performing experiments to probe the properties of a wide range of materials. Usually, the experimental procedures involve extensive surface preparation using sputtering and annealing and are performed in highly controlled environments. Tip preparation and characterisation is also crucial and consists of a substantial part of the preparation of the experiment. This is a vast subject and can get very specific to the material studied. There exist many reviews and monographs in the subject, two of which are by Roland Wiesendanger [37] and Foster and Hofer [4].

3.11 Bibliography

[1] G. Binnig and H. Rohrer. Scanning Tunneling Microscope. *Helvetica Physica Acta*, 1982.

[2] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel. Surface Studies by Scanning Tunneling Microscopy. *Physical Review Letters*, 49(1):57–61, 1982.

[3] H. Rohrer. Scanning Tunneling Microscopy: A Surface Science Tool and Beyond. *Surface Science*, 299-300:956–964, 1994.

[4] Adam S. Foster and Werner A. Hofer. *Scanning Probe Microscopy. Atomic Scale Engineering by Forces and Currents.* Springer, 2006.

[5] Gerd Binnig and Heinrich Rocher. Scanning Tunneling Microscopy - From Birth to Adolescence. *Reviews of Modern Physics*, 59(3):615–625, 1987.

[6] A. S. Foster, W. A. Hofer, and A. L. Shluger. Quantitative Modelling in Scanning Probe Microscopy. *Current Opinion in Solid State and Materials Science*, 5(5):427–434, 2001.

[7] J. Bardeen. Tunneling from a Many-Particle Point of View. *Physical Review Letters*, 6(2):57–59, 1961.

[8] J. Tersoff and D. R. Hamann. Theory of the Scanning Tunneling Microscope. *Physical Review B*, 31(2):805–813, 1985.

[9] J. Tersoff and D. R. Hamann. Theory and Application of the Scanning Tunneling Microscope. *Physical Review Letters*, 50(25):1998–2001, 1983.

[10] Lewis H Ryder. *Quantum Field Theory.* Cambridge University Press, 1996.

[11] Henrik Bruus and Karsten Flensberg. *Many-Body Quantum Theory in Condensed Matter Physics.* Oxford University Press, 2004.

[12] Leo Kadanoff and Gordon Baym. *Quantum Statistical Mechanics.* Westview Press, 1994.

[13] Jorgen Rammer. *Quantum Field Theory of Non-Equilibrium States.* Cambridge University Press, 2007.

[14] Murray Gell-Mann and Francis Low. Bound States in Quantum Field Theory. *Physical Review*, 84(2):350, 1951.

[15] V. B. Berestetskii, L. P. Pitaevskii, and E.M. Lifshitz. *Introduction to Quantum Field Theory*. Butterworth-Heinemann, 1982.

[16] Michael E. Peskin and Dan V. Schroeder. *Course On Theoretical Physics: Quantum Electrodynamics*. Westview Press, 1995.

[17] J Schwinger. Brownian Motion of a Quantum Oscillator. *Journal of Mathematical Physics*, 2(3):407, 1961.

[18] A G Hall. Non-Equilibrium Green's Functions: Generalized Wick's Theorem and Diagrammatic Perturbation Theory with Initial Correlations. *Journal of Physics A: Mathematical and General*, 8(2):214, 1975.

[19] L V Keldysh. Diagram Technique for Nonequilibrium Processes. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, 47:1515–1527, 1964.

[20] D C Langreth. Linear and Non-linear Electron Transport in Solids. *NATO Advanced Study Institute Series*, 17:3, 1976.

[21] H. Haug and A. P. Jauho. *Quantum Kinetics in Transport and Optics of Semiconductors*. Springer Solid-State Sciences, 1996.

[22] C Caroli, R Combescot, P Nozieres, and D Saint-James. A Direct Calculation of the Tunneling Current IV: Electron-Phonon Interaction Effects. *Journal of Physics: Solid State Physics*, 5(1):21, 1971.

[23] Yigal Meir and Ned S. Wingreen. Landauer Formula for the Current through an Interacting Electron Region. *Physical Review Letters*, 68(16):25122515, 1992.

[24] R Landauer. Spatial Variation of Currents and Fields due to Localised Scatterers in Metallic Conduction. *IBM Journal of Research and Development*, 1(3):223–231, 1957.

[25] M Büttiker et al. . Generalized Many-Channel Conductance Formula With Application to Small Rings. *Physical Review B*, 31(10):62076215, 1985.

[26] Antti-Pekka Jauho, Ned S. Wingreen, and Yigal Meir. Time-Dependent Transport in Interacting and Noninteracting Resonant-Tunneling Systems. *Physical Review B*, 50(8):5528, 1994.

[27] A A Abrikosov et al. . *Methods of Quantum Field Theory in Statistical Physics*. Dover Publications Inc., 1976.

[28] Kurt Stokbro et al. TranSIESTA: A Spice for Molecular Electronics. *Annals of the New York Academy of Sciences*, 1006, 2006.

[29] T. E. Feuchtwang. Tunneling Theory Without the Transfer-Hamiltonian Formalism. IV. The Abrupt (Zero-Width) Three-Dimensional Junction. *Physical Review B*, 13(2):517530, 1976.

[30] J B Pendry, A B Pretre, and B C H Krutzen. Theory of the Scanning Tunneling Microscope. *Journal of Physics: Condensed Matter*, 3(24):4313, 1991.

[31] B. C. Stipe, M. A. Rezaei, and W. Ho. Coupling of Vibrational Excitation to the Rotational Motion of a Single Adsorbed Molecule. *Physical Review Letters*, 81(6):12631266, 1998.

[32] B. C. Stipe, M. A. Rezaei, and W. Ho. Single-Molecule Vibrational Spectroscopy and Microscopy. *Science*, 280(5370):1732–1735, 1998.

[33] Michael Galperin and Mark A. Ratner and Abraham Nitzam. Inelastic Electron Tunneling Spectroscopy in Molecular Junctions: Peaks and Dips. *Journal of Chemical Physics*, 121(23):11965, 2004.

[34] N Lorente and M Persson. Theory of Single Molecule Vibrational Spectroscopy and Microscopy. *Physical Review Letters*, 85(14):29973000, 2000.

[35] N Lorente. Mode Excitation Induced by the Scanning Tunneling Microscope. *Applied Physics A*, 78(6):799–806, 2004.

[36] M. L. Bocquet1, H. Lesnard1, and N. Lorente. Inelastic Spectroscopy Identification of STM-Induced Benzene Dehydrogenation. *Physical Review Letters*, 96(9):096101, 2006.

[37] Roland Wiesendanger. *Scanning Probe Microscopy and Spectroscopy: Methods and Applications*. Cambridge University Press, 1994.

Chapter 4

Simple Defects on Metal-Oxide Nanotubes

4.1 Introduction

Since the discovery of carbon nanotubes by Ijima in 1991 [1], the field of carbon-based and inorganic nanotube materials has grown considerably due to their potential applications in electronics, photonics, chemical separation, (photo)catalysis and biotechnology [2–6]. While carbon-nanotubes and their inorganic counterparts such as BN, WS_2 and MoS_2 are routinely produced by electric-arc discharge, chemical vapor deposition, or laser ablation, new low-temperature solution-phase chemistry routes have recently been proposed for semiconducting single-walled aluminosilicate (Al-Si) and aluminogermanate (Al-Ge) nanotubes [7, 8]. The reported in [7], [8] extremely high experimental control in terms of dimensions and monodispersity of these nanostructures, together with the potentially huge range of tunable properties via chemical functionalization and substitutional doping [9], make both Al-Ge and Al-Si attractive candidates as large-storage chemical devices [7, 8], artificial ion-channel systems, and insulating coatings for conducting cores. This could make it possible to enforce one-dimensional anisotropic conductivity at the nanoscale [2–6, 10]. Al-Si and Al-Ge nanotubes are structurally analogous to the naturally occurring hydrous-aluminosilicate Imogolite found in volcanic soils [11, 12].

The walls of Imogolite consist of a single layer of octahedrally coordinated aluminium hydroxide (gibbsite), with tetrahedral silanol (Si-OH) groups attached at the inner side

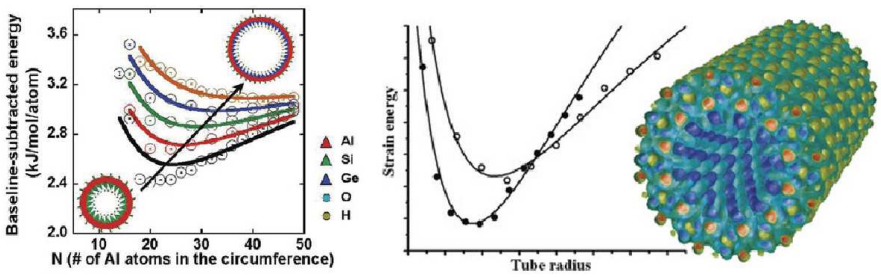

Figure 4.1: Left: Total energy per atom for AlSi(Ge)OH nanotubes as a function of the number of Al atoms in the nanotube circumference [13]. Right: Calculated strain energy as a function of the radius R for zigzag and armchair Imogolite nanotubes [14].

of the tube as in Fig. 4.2. From a compositional point of view, the only difference between Al-Si and Al-Ge tubes is the substitution of silanol groups with germanol (Ge-OH) moieties. The resulting chemical formula of the unit cell is $(Al_2SiO_7H_4)_N$ and $(Al_2GeO_7H_4)_N$ for the Al-Si and Al-Ge tubes, respectively. N refers to the number of radially nonequivalent aluminum atoms along the nanotube circumference. N is necessarily an even number for symmetry reasons [15]. Regardless of the specific route adopted for their synthesis, both Al-Si and Al-Ge tubes are achiral and analogous to zig-zag $(n, 0)$ semiconducting carbon-based nanotubes [16]. A recent Density Functional Theory (DFT) study of Imogolite-based nanotubes performed by Alvarez-Ramirez [17] opened up the possibility of first-principle studies for these systems. Similarly to previous simulations of neutral paramagnetic defects in zeolites [18], the effects of the simplest defect, the neutral hydroxyl (-OH) vacancy, on Al-Ge and Al-Si is addressed in this chapter.

4.2 Methodology

The main part of the first-principles study concerning the AlSi(Ge)OH nanotubes themselves, was performed using Density Functional Theory with a plane-wave basis set as implemented in the code VASP [19, 20]. The plane-wave energy cut-off was set to 400 eV. The atoms were represented using the Projector Augmented Wave method [21, 22] while exchange and correlation were approximated using the GGA-PW91 exchange and correlation functional [23]. The sampling of the irreducible Brillouin zone was initially performed on the Γ point only, however a 2 special k-point grid was eventually adopted [24]. For the geometry optimisations, atoms were relaxed until all spatial components of all inter-atomic forces were less than 0.03eV/Å. The 2-dimensional equivalent sheets

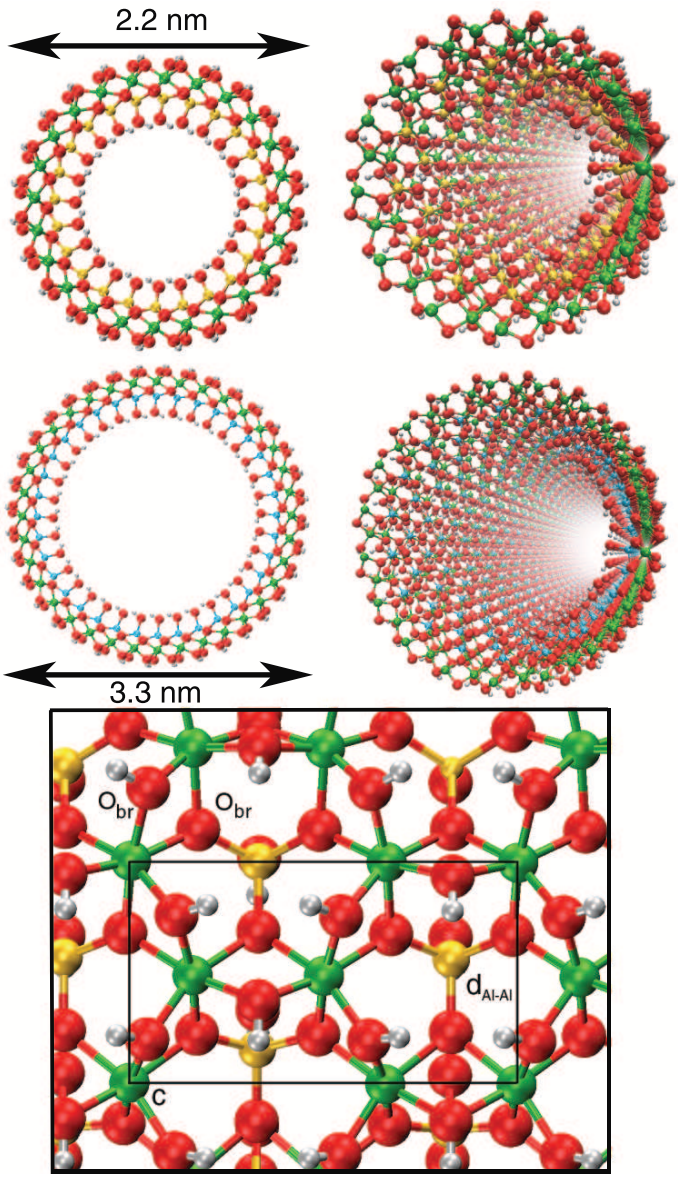

Figure 4.2: Optimised geometrical structure of $(Al_2SiO_7H_4)_{24}$ (top) and $(Al_2GeO_7H_4)_{36}$ (middle) based nanotubes. The single-wall structural motif (bottom) is displayed together with the zig-zag periodic unit of size $\{c, d_{Al-Al}\}$ along the nanotube axis and circumference. O: red, H: gray, Al: green, Si: yellow, Ge: cyan.

were also studied for comparison purposes using basis sets consisting of Gaussian functions[1]. Both pure GGA-PW91 and hybrid B3LYP [26] approximations were adopted for exchange and correlation as implemented in the code Gaussian [27]. Hydroxyl vacancies OH_{vac} were modeled by removing a neutral hydroxyl fragment from either the inner or the outer surface of the nanotube's wall. For these calculations a spin-polarized approach was adopted. Defect formation energies were calculated as total-energy differences according to the equation below

$$E_{\mathrm{DFE}} = E_{\mathrm{Tube}} - E_{\mathrm{OHvac}} - E_{\mathrm{OH(g)}}. \tag{4.2.1}$$

E_{Tube}, E_{OHvac} and $E_{\mathrm{OH(g)}}$ are the total energies of the pristine tube, the reduced tube, and the single hydroxyl in the gas phase respectively.

4.3 Geometrical Optimization

The nanotubes were constructed assuming an octahedral coordination for Al atoms and a tetrahedral coordination for Si and Ge atoms. No distortions were introduced and the bond lengths used were consistent with the optimized CLAYFF4 nanotube radius of curvature [15]. Because of the increased computational requirements for the first-principles DFT studies of the kind performed in this case, the number of Si and Ge atoms in the unit cell was limited to 24 and 36 respectively. The chemical formulas for the two nanotubes is given, in this case, by $(Al_2SiO_7H_4)_{24}$ and $(Al_2GeO_7H_4)_{36}$. The total number of atoms is 336 and 504 for the simulations cells in the two cases respectively. Although these numbers are consistent with other experimental and theoretical studies [7, 8, 14], the number of gibbsite-like units contradicts the values reported in [17]. In order to determine the optimal length of the unit cell along the nanotube axis c, the size of the repeat unit was scanned with the geometries optimised at every step. The value of the total energy of the nanotube per Al atoms lying in its circumference was then plotted against the length, and the data were fitted using the least square non-linear fitting method as in Fig. 4.3. The values obtained that minimized the total energy were 8.68±0.01Å and 8.78±0.01Å for Al-Si and Al-Ge respectively. These values seem slightly overestimated compared to X-ray diffraction experimental values obtained for Al-Si and

[1]The basis set used is the CEP-121G [25]. It yielded results in agreement with previous band-gap calculations performed with PW91 and a double-zeta basis set with polarization functions, 7.0 eV versus 6.8 eV in the latter case [18].

Al-Ge thin films, reported to be 8.51Å in [7, 8]. Besides the tendency of GGA-PW91 to overestimate aluminosilicate bond lengths, this discrepancy can also be attributed to interactions of the tubes with the crystallization solvent, or to interactions among the nanotubes themselves when present in thin films. These nuances were not considered in this theoretical study. Since the values reported in [17] for Al-Si with N = 20 are 8.78Å, 8.72Å or 8.62Å, depending on the basis set and the size of the k-point grid used for the BZ sampling, it can be suggested that the size of the repeat cell along the nanotube axis depends only weakly on the number of gibbsite-like units present.

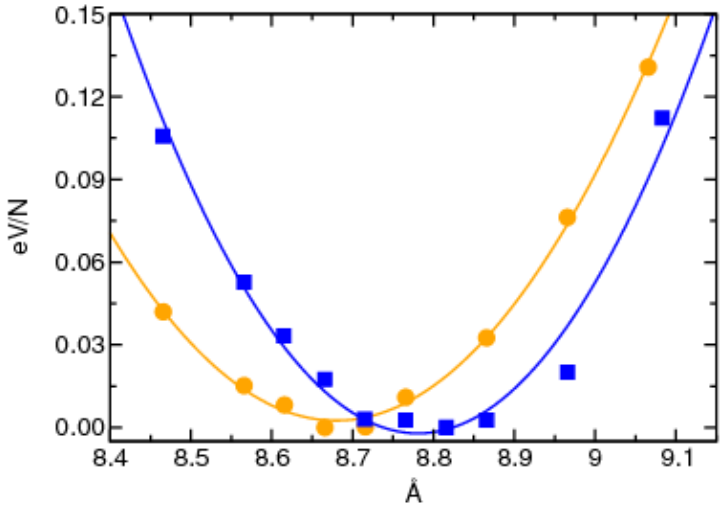

Figure 4.3: Total nanotube energy per number of Al atoms along the nanotube circumference (N) versus length of the repeat unit along the tube axis (c). Al-Ge tube: blue, Al-Si tube: yellow.

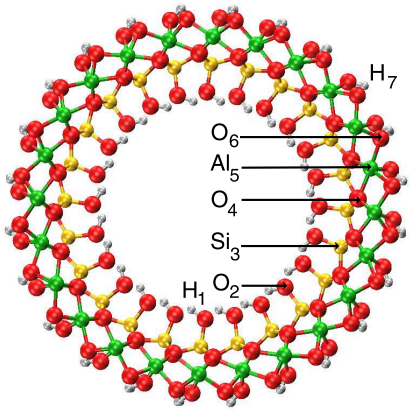

Figure 4.4: Radial atomic labeling for Al-Si (Al-Ge). O: red, H: gray, Al: green, Si: yellow.

For comparison purposes, the corresponding 2-dimensional sheets of the aluminosilicate and aluminogermanate nanotubes were also considered. The process of determining the

Table 4.1: Atomic radial distribution for Al-Si optimized nanotubes. All diameters are reported in Angstroms.

	Al-Si	Al-Si-OH_{invac}	Al-Si-OH_{outvac}
H_1	12.33 ± 0.01	12.32 ± 0.04	12.31 ± 0.08
O_2	13.27 ± 0.02	13.27 ± 0.03	13.25 ± 0.08
Si_3	16.52 ± 0.02	16.52 ± 0.02	16.51 ± 0.09
O_4	17.82 ± 0.07	17.81 ± 0.07	17.79 ± 0.10
Al_5	19.75 ± 0.03	19.74 ± 0.04	19.71 ± 0.20
O_6	21.91 ± 0.06	21.91 ± 0.06	21.89 ± 0.09
H_7	23.15 ± 0.05	23.15 ± 0.05	23.15 ± 0.09

optimal geometry for the sheets followed a similar to the nanotubes methodology as both the size and the angle of the unit cell were varied. For those values that yielded the lowest total energies, the geometries were optimized, giving the following optimal values for the z-size of the unit cell (c) and the distance between aluminum atoms (d_{Al-Al}): $\{8.565, 4.945\}$Å and $\{8.652, 4.995\}$Å for Al-Si and Al-Ge respectively. Corresponding values for Al-Si obtained using a double zeta basis set with polarization orbitals (DZP), $\{8.46, 4.9\}$Å [17], were in good agreement with this study. The bending energies were evaluated to be 0.45eV/N and 0.19eV/N for Al-Si and Al-Ge respectively, suggesting that the tube bending is a consequence of minimizing the strain due to the mismatch between the stronger Si(Ge)-O and weaker Al-O bonds. The more bent Al-Si is getting more stabilized compared to Al-Ge by the bending process.

Atomic radial distributions for Al-Si and Al-Ge optimized nanotubes with and without the presence of the defect OH_{vac} are reported in Tables 4.1 and 4.2 respectively. The labeling follows the atomic labeling in Fig. 4.4. The calculated inner and outer diameters were 12.33±0.01Å and 22.09±0.03Å for Al-Si and 23.15±0.05Å and 33.15±0.04Å for Al-Ge respectively. The radial distribution of the atomic layers with $O_2 - Si_3$ ($O_4 - Al_5$) separations larger than $Si_3 - O_4$ ($Al_5 - O_6$) distances is consistent with the reported structural data for Imogolite crystals [11, 12, 28]. By comparing the Al-Si and the Al-Ge nanotubes, it can be seen that the distance between $O_2 - Ge_3$ is again larger than $Ge_3 - O_4$. The difference between the distances $O_4 - Al_5$ and $Al_5 - O_6$, however, are modeled to be very close, with a deviation of less than 0.1Å. Defect formation energies are reported in Table 4.3. Although the values suggest relatively large exothermic formation energies, the presence of the OH_{vac} defect does not alter significantly the tube radii and monodispersity. The localised distortions introduced by the removal of the hydroxyl group induce maximum deviations to the bond lengths of less than 0.1Å with

Table 4.2: Atomic radial distribution for Al-Ge optimized nanotubes. All diameters are reported in Angstroms.

	Al-Ge	Al-Ge-OH_{invac}	Al-Ge-OH_{outvac}
H_1	22.09 ± 0.03	22.07 ± 0.05	22.06 ± 0.08
O_2	22.94 ± 0.03	22.93 ± 0.02	22.88 ± 0.12
Ge_3	26.43 ± 0.02	26.42 ± 0.09	26.38 ± 0.11
O_4	27.78 ± 0.04	27.75 ± 0.06	27.73 ± 0.12
Al_5	29.75 ± 0.02	29.73 ± 0.02	29.69 ± 0.20
O_6	31.90 ± 0.04	31.89 ± 0.05	31.86 ± 0.09
H_7	33.15 ± 0.04	33.14 ± 0.05	33.12 ± 0.12

Table 4.3: Defect formation energies. Values calculated as total energy differences.

Al-Si-OH_{invac}	Al-Si-OH_{outvac}	Al-Ge-OH_{invac}	Al-Ge-OH_{outvac}
+6.08 eV	+6.32 eV	+4.51 eV	+5.71 eV

respect to the averaged radial atomic distribution. This makes them hard to detect even by Transmission Electron Microscopy experiments like those reported in [7, 8].

Structural parameters for Al-Si and Al-Ge optimized nanotubes and the corresponding 2D sheets are reported in Table 4.4. Upon inspection of the averaged bond lengths and angles for the optimised systems, the Al-O bond lengths are found to be in good agreement with previous GGA calculations on aluminosilicate clusters. They are practically not affected by the presence of either silanol or germanol groups in the tube cavity. The changes at Al-O level are minimal also with respect to values for the analogous 2D-sheet. The same applies also to O-H bond lengths. The main changes arising from the tube bending can be found in the Si(Ge)-O bonds, as indicated by the larger variations compared to the values of the corresponding 2D-sheets. With a shorter Si(Ge)-O bond length with respect to Al-O, the simulations confirm the bending process as a consequence of minimizing the structural strain originating from the stronger Si(Ge)-O bonds on the inner surface and the weaker Al-O bonds on the outer wall of the tubes. Finally, with a maximum deviation of $\sim 12^\circ$ from an ideal octahedral (90°) coordination, the optimised bending angles indirectly support the assumption made in [15] that the Al-O coordination environment can be approximated by a rigid Al-O octahedron. However, it should be stressed that the present values have been calculated disregarding any solvent effects which may play a significant role in the bending process.

Although the presence of one OH_{vac} (both inside and outside the tube cavity) is calculated to marginally perturb the global tube monodispersity, on a local scale the defects induce different geometrical distortions depending on the specific metallic atom were

Table 4.4: Structural parameters for Al-Si and Al-Ge optimized nanotubes. The data for the analogous 2D-sheets are reported within brackets.

	Al-Si	Al-Ge
Al-O (Å)	1.918 ± 0.052 (1.922 ± 0.072)	1.916 ± 0.054 (1.917 ± 0.057)
Si(Ge)-O (Å)	1.645 ± 0.005 (1.657 ± 0.028)	1.773 ± 0.013 (1.783 ± 0.023)
O-H (Å)	0.970 ± 0.010 (0.871 ± 0.009)	0.969 ± 0.011 (0.972 ± 0.008)
O_{br}-Al-O_{br} (°)	78.6 ± 1.4 (82.9 ± 4.2)	78.0 ± 1.6 (81.7 ± 2.7)

OH_{vac} is created. In fact, while the Si(Ge)-O bonds increase from 1.645 (1.773) Å to 1.662 (1.817) Å, one OH_{outvac} results in a reduction of 0.027 Å and 0.052 Å for the Al-O bond opposite to the vacancy for Al-Si and Al-Ge, respectively. Again, the increased distortions for the Al-Ge tube can be considered as another indication of the higher thermodynamic stability of the Al-Si bond compared to the Al-Ge bond. Concerning the changes on the bending angles induced by one OH_{vac}, the vacancies are calculated to cause a reduction of the O-Al-O angle inside the tube walls to 77.4° (77.2°) for $Al - Si(Ge) - OH_{invac}$ and an increase to 80.8° (80.3°) for $Al - Si(Ge) - OH_{outvac}$.

4.4 The Electronic Structure of AlSiOH and AlGeOH Nanotubes

The calculated Density of States for the Al-Ge, Al-Si nanotubes and the corresponding 2D sheets are reported in Fig. 4.5. The band gaps were found to be 4.1 eV and 3.9 eV for Al-Si and Al-Ge respectively. These values are in agreement with simulations of aluminosilicate clusters (∼4 eV [29]) and aluminum oxide thin films (∼4 eV [30]) obtained using the same PW91 functional. They also agree well with other recent DFT studies on $N = 20$ Al-Si and Al-Ge nanotubes [17]. When considering the Density of States of the defect-free nanotubes along with their equivalent 2-dimensional sheets, it can be noticed that the physical bending does not affect substantially the less wrapped Al-Ge tube. This is in contrast to Al-Si in which the difference in the calculated band-gap value between the tube and the sheet is 0.4 eV. The same value fir Al-Ge is 0.1 eV.

Due to the constraints in the available computational resources, the hybrid DFT calculations with the B3LYP functional were carried only with the 2D sheets. There, the band gaps where found to be overestimated for both systems. Despite previous studies [31] in which B3LYP was successfully used for the study of the chemical reactivity of

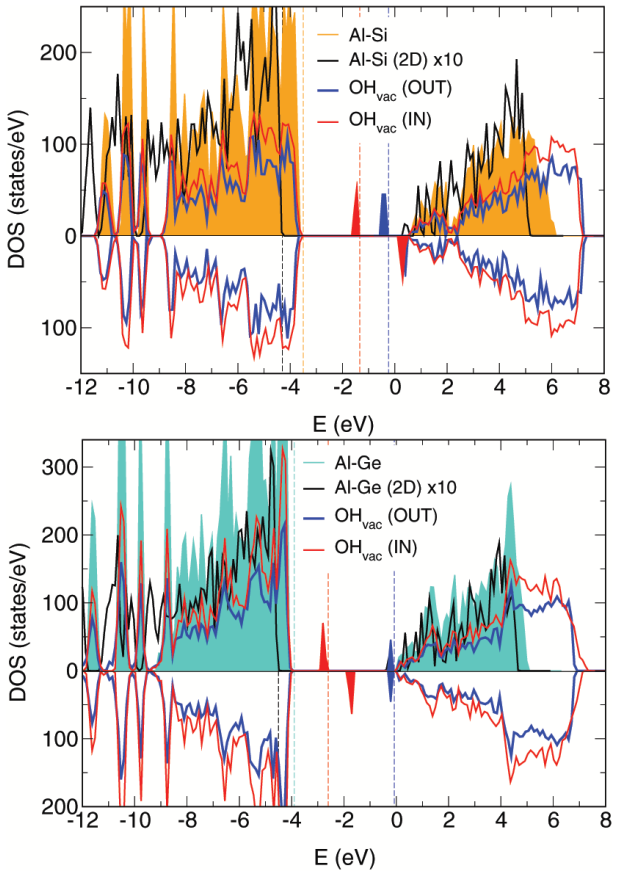

Figure 4.5: Total Density of States (DOS) for defect-free Al-Si and Al-Ge, their 2D analogues, and in the presence of one OH_{vac} both on the outer (OUT) and the inner (IN) surface of the tubes. Calculated Fermi energies are displayed as dotted lines with the same color labeling as for the DOS. 2D and band gap defect states (filled) have been increased by a factor 10 for clarity.

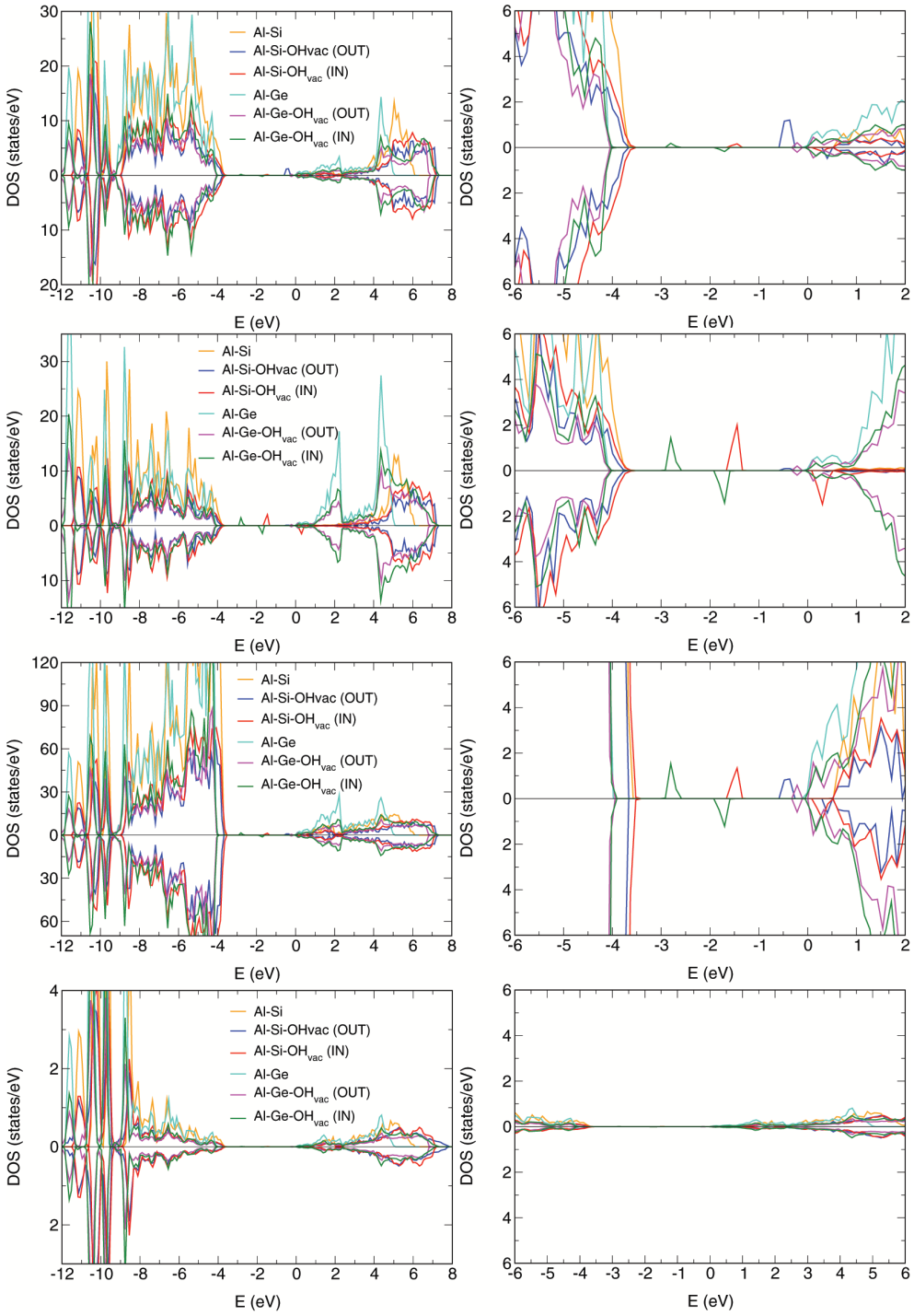

Figure 4.6: Atom resolved Projected Density of States (PDOS) for optimized aluminosilicate and aluminogermanate nanotubes (left) and corresponding close-up of the band gap states (right). From top to bottom: Al, Si(Ge), O, H. The contribution from Al-Ge tubes has been scaled down by a factor of 1.5 for displaying purposes.

Zeolites, it was calculated that the band-gaps were 9.3 eV for Al-Si and 7.42 eV for Al-Ge respectively. This is in contrast to the experimental value of 3.6 eV for the optical gap. Although this discrepancy is quite substantial, this overestimation is in agreement with previous results obtained using Hartree-Fock methods [18] in which calculated band gaps were wider than those obtained using pure DFT with GGA-PW91 for exchange and correlation. It is then expected that, although no explicit calculations were performed with the B3LYP functional on the nanotubes themselves, and because of the relatively small effect of the bending on the band-gap, the values for the band-gap should be in the range $[4.1 - 9.3]$ eV for Al-Si and $[3.9 - 7.4]$ eV for Al-Ge. These upper limits agree relatively well with the values calculated using tight binding DFT which are known to be overestimated [14].

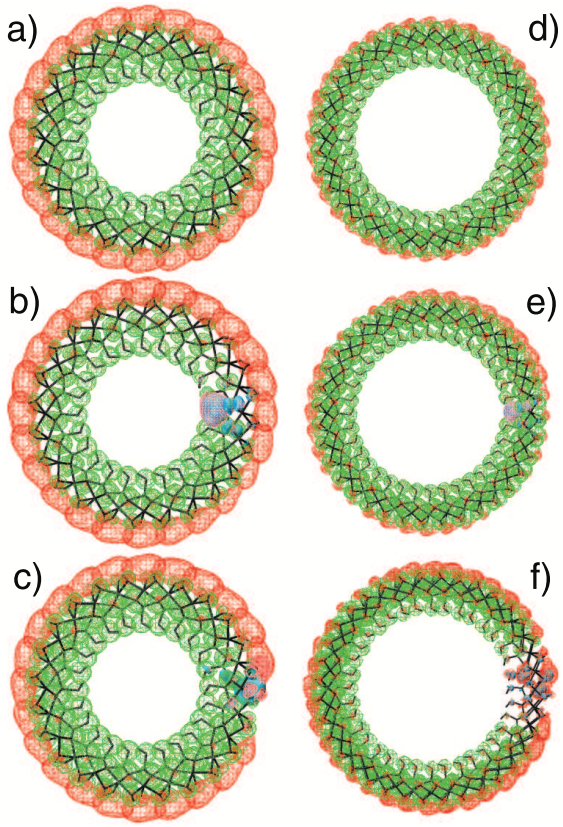

Figure 4.7: Band-decomposed charge density for defect-free Al-Si (a) and Al-Ge (d), and in the presence of one OH_{vac} both inside (b,e) and outside (c,f) the nanotube cavity. The VB (green, 10^{-6}eÅ^{-3}) and CB (red, 5×10^{-7}eÅ^{-3}) densities have been integrated over 0.5 eV from the band onset. Occupied and empty band-gap defect states are displayed (5×10^{-7}eÅ^{-3}) in cyan and pink, respectively.

Band-decomposed charge densities for both the valence band (VB) and the conduction band (CB) edges, as can be seen in Fig 4.7, were calculated by considering electronic states within 0.5 eV from the band onsets. It can be noticed that both Al-Si and Al-Ge are characterized by a precise real-space separation of VB and CB. According to the

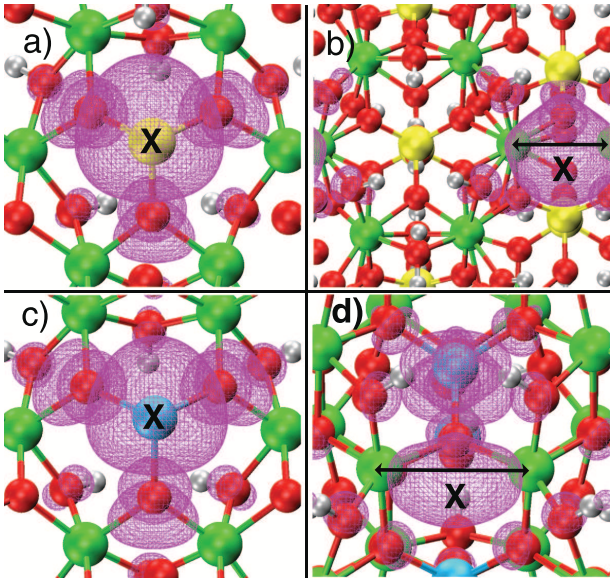

Figure 4.8: Close up of band decomposed charge density (5×10^{-7}eÅ^{-3}) for occupied defect states associated with OH_{vac} (**X**). a: $Al-Si-OH_{invac}$; b: $Al-Si-OH_{outvac}$; c: $Al-Ge-OH_{invac}$; d: $Al-Ge-OH_{outvac}$. OH_{invac} and OH_{outvac} are shown from inside and outside the tube cavity, respectively. Under-coordinated Al atoms for OH_{outvac} are highlighted by mean of arrows. O: red, H: gray, Al: green, Si: yellow, Ge: cyan.

decomposition, the valence band edge is localised inside the nanotube cavity, while the conduction band edge faces the outer side of the nanotube. Atom resolved Projected Density of States (PDOS) for the optimised aluminosilicate and aluminogermanate nanotubes are reported in Fig 4.6. The plot reveals a predominant oxygen contribution to both the high-energy valence band edge of $2p$ character and to the low energy conduction band edge ($2s$-$2p$ hybridized states), regardless of the presence of OH_{vac}. This is in agreement with [14, 17] for the VB edge however the predominant component of the CB edge is assigned to O atoms and not to (outer) H_7 (Fig. 4.4) atoms as in [17] or $3p$ and $3d$ Si states and $3s$ Al states as suggested in [14]. This discrepancy can be attributed to the small energy cut-off of $10eV$ imposed in the PDOS analysis in [17] and to the minimal atomic basis set used in [17] and [14]. The modeled real space localisation of both the VB and CB edges suggests a strong radial anisotropic electron affinity with an ensuing enhanced Brønsted acidity (basicity)[2] for the inner (outer) tube surfaces as proposed in [28] and implied by simulations in [14, 17].

In order for the effects of OH vacancies on the global electronic structure of the tube to be investigated, one hydroxyl fragment was eliminated both on the inner, $Al-Si(Ge)-OH_{invac}$, and on the outer, $Al-Si(Ge)-OH_{outvac}$, side of the nanotube wall. In analogy with O vacancies on other metal-oxide substrates [32], the presence of one OH_{vac} is found to

[2] A Brønsted acid is a proton donor. Correspondingly, a Brønsted base is a proton acceptor.

introduce electronic states in the pristine band gap as it can be seen in Fig. 4.5. One OH_{invac} is found to create both occupied and unoccupied defect states, which reduce the actual band-gap to ∼1.8 eV and ∼1.1 eV for both Al-Si and Al-Ge. This effect results in a substantial enhancement of the conductivity of the global systems. Both defects were modeled to be magnetic, generating a doublet spin-state. As with previous results for paramagnetic defects in zeolites [18], the occupied (unoccupied) defect states for both for Al-Ge and Al-Si are highly localised around the under-coordinated Si (Ge) atom (Fig. 4.7, 4.8). It should be noted, however, that while one OH_{vac} on the outer wall of Al-Si is also found to create a localised paramagnetic (doublet) spin-state (Fig. 4.7, 4.8) resulting in a reduction of the actual band gap (0.8 eV), the same defect in the Al-Ge analogue forms a state just at the CB onset which pins the Fermi level there. Consequently, despite the odd total electron count in the system, the DFT energy is minimized by a non-magnetic solution with equal occupancy of both up and down spins (Fig. 4.5 and 4.6).

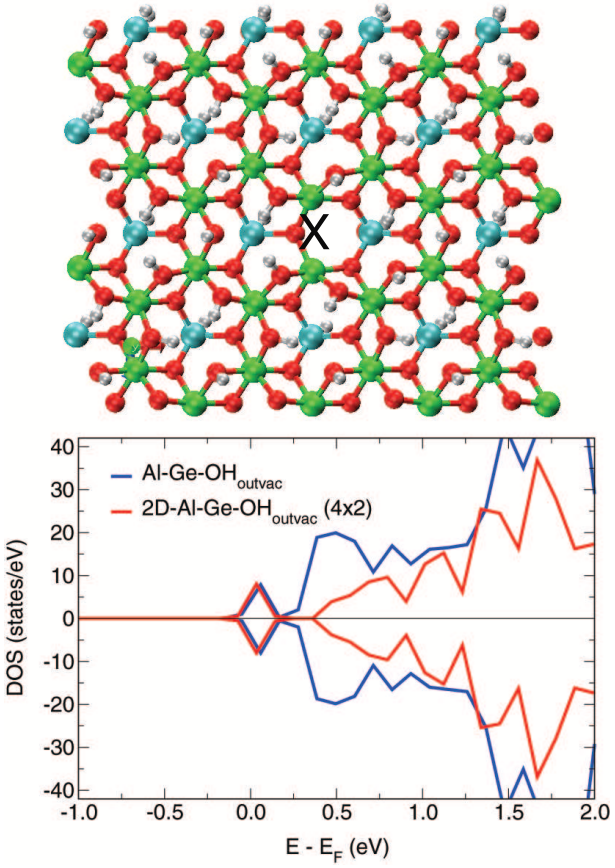

Figure 4.9: Top: 2-Dimensional Al-Ge 4x2 super-cell with a hydroxyl vacancy on the outer surface. The X mark indicates the position of the hydroxyl vacancy. Bottom: Close-up of the onset of the sheet's Conduction Band in comparison with the CB of the equivalent Al-Ge nanotube. O: red, H: gray, Al: green, Ge: cyan.

This behaviour has raised the question on whether it is real and whether it could also be found on a truly isolated defect or is it an artifact of the imposed periodic boundary conditions. In order for this to be assessed, one OH_{vac} was introduced on the outer topmost face of a 4x2 super cell of the 2-dimensional Al-Ge sheet. This increased the distance between defects from 8.78Å in $Al-Si(Ge)-OH_{outvac}$ nanotubes to 17.3Å in the 2D sheet. Also in this case, the defect was modeled to introduce a non-magnetic solution suggesting that the cause of this metallization effect is the smaller size of the band gap of Al-Ge with respect to Al-Si and not a computational artifact. Fig. 4.9 showcases the small shift of the band-gap states after increasing the size of the simulation cell for the 2-dimensional sheet.

4.5 Surface Polarisation

It has been suggested that surface charge properties and chemical reactivity of Imogolite depend on a delicate (pH-dependent) balance between local geometrical deformations, associated with the tube bending, and positive (negative) charge accumulation on the outer (inner) surface [28]. Charge separation across the tube walls is expected to play a fundamental role in the adoption of Al-Si (Al-Ge) as insulating coating around conducting cores [7, 8, 10]. In this respect, it is of interest to assess the changes induced by OH_{vac} in terms of both local and global polarization across the tube section.

By using Gauss's theorem within a cylindrical condenser approximation [33], it is possible to relate the change ($\Delta\overline{V}$) in the (microscopically) averaged electrostatic potential across the polarised interface to the electrostatic dipole density across the interface itself (μ_σ) as follows:

$$\Delta\overline{V} = 4\pi\mu_\sigma \ln\left(\frac{R_{in}}{R_{out}}\right) \tag{4.5.1}$$

Atomic units have been used and R_{in} (R_{out}) defines the inner (outer) radius of the tube. The microscopically averaged potential $\overline{V}$ in cylindrical coordinates ($r = \sqrt{x^2+y^2+z^2}, x = r\cos\phi, y = r\sin\phi, z$) reads:

$$\overline{V}(r) = \frac{1}{2\pi c_0}\int\int V(r,\phi,z)\,dz\,d\phi \tag{4.5.2}$$

where c_0 defines the optimised length of the simulation cell along the tube axis z (Fig. 4.2).

It is possible for the actual dipole density to be evaluated across the tube walls from the $\Delta\overline{V}$ values calculated from the DFT simulations. Besides providing the electrostatic potential (V), DFT simulations can also be conveniently used to infer a value for R_{in} (R_{out}) which accounts not only for atomic positions but also for the electronic distribution of the nanotube itself. From the onset of the electrostatic plateaus characterizing the vacuum region (Fig. 4.10), it is also possible to define the spatial extent of the total charge distribution of the tube. Alternatively, one could define R_{in} (R_{out}) on the basis of the real space positions for which the microscopically averaged (global) charge density ($\overline{\rho}$) becomes zero within the convergence tolerance enforced in the DFT simulation itself. Since the electrostatic potential and the (global) charge density are related by Poisson's equation [$\nabla^2\overline{V}(x,y,z) = -4\pi\overline{\rho}(x,y,z)$], the two approaches necessarily yield equivalent results as shown in Fig. 4.10. It can be noted that, according to this procedure, the calculated polarisation across the tube walls accounts for the total distribution of the tubes charge density. In addition, the contribution from interstitial charge and local gradients also contributes to the final net polarization. Moreover, the converged plane-wave approach yields results independent of the basis set, at odds with Mulliken based analysis [14, 17]).

In the present analysis, the electrostatic potential, V, was evaluated as given by the VASP program [19, 20, 22] and the global charge density ρ was calculated by the Fast Fourier Transform (FFT) expansion of the Poisson's equation in reciprocal space. This was achieved by calculating first the FFT coefficients $\hat{V}_G$ of the electrostatic potential, using the same grid ($G = 0, 1, 2,... G_{max}$) as used in the corresponding VASP simulations. The FFT coefficients of the charge density were obtained as $\hat{\rho}_0 = 0$, $\hat{\rho}_G = \frac{G^2}{4\pi}\hat{V}_G$, where G defines the length of the reciprocal lattice vectors used in the FFT expansion. Finally, ρ was obtained by back transformation of $\hat{\rho}_G$ [34]. FFT evaluations were performed via the open source (complex double precision) netlib FFT-pack [35]. From the change in the averaged electrostatic potential, Al-Ge (R_{in}=8.7 Å, R_{out}=18.2 Å) is calculated to possess a dipole density across the wall (0.06 DebyeÅ^{-2}), which is twice the value for Al-Si (0.03 DebyeÅ^{-2}, R_{in}=4.2 Å, R_{out}=13.7 Å). Thus, in agreement with [28] and [14] (but at odds with [17]), the calculated polarisation directions agree with the positive (negative) charge accumulation at the outer (inner) surface as suggested by the higher inner $\overline{V}$ plateau with respect to the outer value for both Al-Si and Al-Ge (Fig. 4.10).

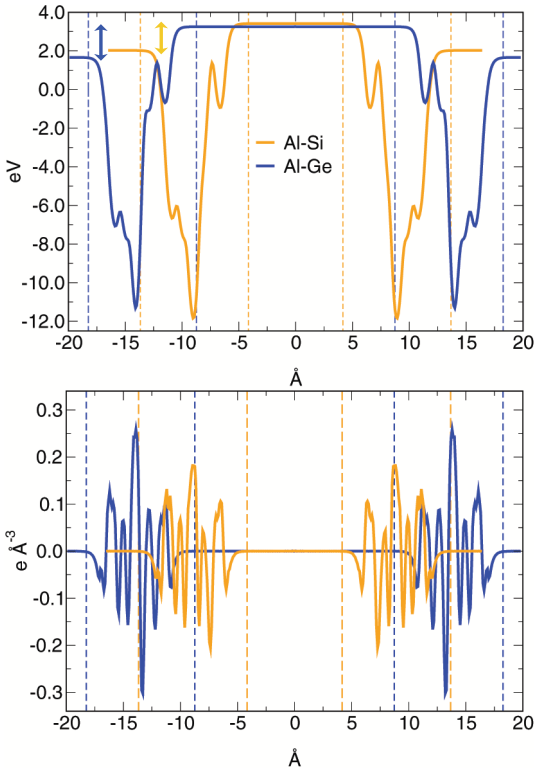

Figure 4.10: Top: radial distribution of the averaged electrostatic potential $\overline{V}$. Vertical arrows highlight the electrostatic potential shifts $\Delta\overline{V}$ across the tube walls. Bottom: radial distribution of the averaged global charge density $\overline{\rho}$. The vertical dotted lines define the radial extent of the tube walls as emerging from the caculated $\overline{V}$ or $\overline{\rho}$ plateaus (see text for details).

With the aim of providing deeper insight into the chemical differences which characterise the outer and inner surfaces of Al-Si and Al-Ge, the change in electrostatic potential associated with specific atomic layers (see Figure 4.4, 4.10 and Table 4.1, 4.2, 4.5) was considered. This kind of analysis is convenient because the results can be directly related to the relative increase/decrease of local Lewis acid-base reactivity for the considered site within each simulation cell. The lower the electrostatic potential, the higher the acidity, i.e. the electron acceptor affinity for the considered site. An O atoms resolved analysis of the electrostatic potential acting at $O_2/O_4/O_6$ level for all the systems considered is reported in Tab. 4.5. While a direct comparison among $O_2/O_4/O_6$ sites of a given system is physically meaningful, on the basis of both the electrostatic averaging procedure enforced in the DFT simulations [34] and the different charge distribution for each simulation cell, no comparison should be attempted between different systems (i.e. different columns of Tab. 4.5) in absence of any electrostatic alignment procedure (which in this case has been neglected).

Disregarding the tube composition and defect presence, the inner O_2 layer is evaluated to be a weaker Lewis acid (i.e. electron acceptor) than the outer O_6 layer as expected on the basis of the experimental [28] and calculated charge distribution across the tube walls (Fig. 4.10 and [14]). The simulations also suggest the O_4 layer as the most acidic site. Due to defect induced charge reorganization, the presence of one OH_{vac} is modeled to alter locally the acid-base properties of the O layers. Specifically, the relative Lewis acidity of O_4/O_6 layers with respect to O_2 increases upon formation of one OH_{invac}, due to the highly localised $Al - Si(Ge) - OH_{invac}$ states (Fig. 4.7, 4.8). Conversely, and not surprisingly, the relative acidity of O_4 with respect to O_2 is reduced when one OH_{outvac} is created on the surface. As a final remark it can be noted that, while convenient in terms of relative Lewis acid-base reactivity, the present analysis for $O_2/O_4/O_6$ layers cannot provide quantitative insights into the global polarization across the nanotube walls. To this end, the global charge organization across the wall, i.e. the inner and outer electrostatic plateaus (Fig. 4.10), must be accounted for and no dipole evaluation should be attempted on the basis of the reduced set of data displayed in Tab. 4.5.

Table 4.5: Averaged electrostatic potential $\overline{V}$ changes (eV) for O_4/O_6 layers with respect to O_2. Angular sectors of $\pm 15^\circ$ ($2\pi/24$) and $\pm 10^\circ$ ($2\pi/36$) around the defect sites have been averaged for $Al-Si-OH_{vac}$ and $Al-Ge-OH_{vac}$, respectively (see Eq. 2).

	Al-Si	Al-Si-OH_{invac}	Al-Si-OH_{outvac}	Al-Ge	Al-Ge-OH_{invac}	Al-Ge-OH_{outvac}
O_2	0.	0.	0.	0.	0.	0.
O_4	-9.8	-11.2	-11.0	-10.6	-10.7	-13.1
O_6	-5.4	-5.5	-5.3	-6.0	-6.4	-1.7

4.6 Mechanical Properties

Due to their possible use as insulating coatings for one-dimensional nanodevices [7, 8, 10], the elastic properties of Imogolite based nanotubes with respect to external mechanical solicitations can also be addressed. This analysis has so far only been investigated at TB-DFT level [14], and no first-principles or experimental results are currently available.

The adopted geometry optimization protocol, provided the required information in order to calculate the simplest mechanical property associated with the considered nanotubes, i.e their Young's modulus (Y). By defining the axial tensile ($\epsilon < 0$) or compressive ($\epsilon < 0$) strain as:

$$\epsilon = \frac{c - c_0}{c_0} \tag{4.6.1}$$

with the optimized length of the simulation cell along the tube axis z referred to as c_0 (Fig. 4.2), the Young's modulus (Y) can be obtained from the computed energies (E) as a function of the tube (simulation cell) length c (Fig. 4.3) [36, 37]:

$$Y = \left(\frac{\partial^2 E}{\partial \epsilon^2} \right)_{\epsilon=0}. \tag{4.6.2}$$

V_0 is the optimised tube volume for a simulation cell of length c_0 while the energy E appearing in Eq. 4.6.2 has been renormalised by the volume V_0. Apart from the computational accuracy of $E(\epsilon)$, a critical issue in Eq. 4.6.2 is the definition of the equilibrium volume V_0. From elementary geometry it follows that for a hollow cylinder $V_0 = 2\pi\, R_{in}\, \Delta R\, c_0$, were R_{in} is the inner radius, ΔR the thickness of the tube wall and c_0 as defined above. It emerges that different choices of R_{in} and ΔR affect the final value of the calculated Young's modulus (Y). Since by definition $\Delta R = R_{out} - R_{in}$, it is possible in this case to rely on the result for the electrostatic potential (or global charge density) analysis (Fig. 4.10) in order to define the values for R_{in} (R_{out}) which account for the electronic distribution of the nanotube. This provides a value for V_0, which is

Table 4.6: Calculated Young's moduli (GPa) for defect free Al-Si and Al-Ge nanotubes. The adopted values for (R_{in}, ΔR) (Å) are shown within brackets (see text for details).

	Al-Si	Al-Ge
$\overline{V}$, $\overline{\rho}$	122 ± 2 $(4.2, 9.5)$	102 ± 4 $(8.7, 9.5)$
H_1, H_2	144 ± 3 $(6.2, 5.4)$	146 ± 7 $(11.1, 5.5)$
O_2, O_6	168 ± 3 $(6.6, 4.3)$	174 ± 8 $(11.5, 4.5)$

representative of the excluded volume at the equilibrium geometry.

The calculated values for Al-Si and Al-Ge tubes are reported in Tab. 4.6 for different choices of R_{in}, R_{out}, determined both on the basis of the modeled $\overline{V}$ plateaus, as it can be seen in Fig. 4.10 and Eq. 2 and optimized (H_1, H_7), (O_2, O_6) radii (see Fig. 4.3 and Tab. 4.1 and 4.2). Similarly with previous DFT-TB suggestions [14], with calculated values in the range 100-200 GPa, the modeled stiffness for Imogolite based system is comparable with GaAs (~ 270 GPa) and MoS_2 (~ 230 GPa) nanotubes but smaller than for other $C_xB_yN_z$ tubes [36–38]. Due to the known underestimation errors of the adopted GGA framework with respect to the elastic properties of extended metal-oxide structures [39], it is expected that these results should be a lower limit for the experimental Young's modulus for both Al-Si and Al-Ge tubes, which has not yet been experimentally determined.

4.7 Conclusions

In this chapter, the presence of hydroxyl vacancies on $(Al_2SiO_7H_4)_{24}$ and $(Al_2GeO_7H_4)_{36}$ single-walled metal-oxide nanotubes was investigated. The induced effects on their geometry, electronic structure and mechanical properties were discussed. It was found that with the exception of one OH_{vac} localised on the outer wall of the $(Al_2GeO_7H_4)_{36}$ tube, these defects induce occupied and empty states in the band gap. Those states were found to be highly localised both in energy as well as in real space. This work consists of the first ab-initio study using Density Functional Theory of these defects and has been published in the Journal of Physics: Condensed Matter [40].

4.8 Bibliography

[1] S. Iijima. Helical Microtubules of Graphite Carbon. *Nature*, 354:56–58, 1991.

[2] Ray H. Baughman, Anvar A. Zakhidov, and Walt A. de Heer. Carbon Nanotubes - The Route Towards Applications. *Science*, 297(5582):787–792, 2000.

[3] M. S. Dresselhaus and H. Dai. Carbon Nanotubes: Continued Innovations and Challenges. *MRS Bulletin*, 29(4):237–243, 2004.

[4] M. Remskar. Inorganic Nanotubes. *Advanced Materials*, 16(17):1497–1504, 2004.

[5] Reshef Tenne and C. N. R. Rao. Inorganic Nanotubes. *Philosophical Transactions of The Royal Society A*, 362(1832):2099–2125, 2004.

[6] Alberto Bianco, Kostas Kostarelos, and Maurizio Prato. Applications Of Carbon Nanotubes in Drug Delivery. *Current Opinion in Chemical Biology*, 9(6):674–679, 2005.

[7] Sanjoy Mukherjee, Veda M. Bartlow, and Sankar Nair. Phenomenology of the Growth of Single-Walled Aluminosilicate and Aluminogermanate Nanotubes of Precise Dimensions. *Chemistry of Materials*, 17(20):4900–4909, 2005.

[8] Sanjoy Mukherjee, Keesuk Kim, and Sankar Nair. Short, Highly Ordered, Single-Walled Mixed-Oxide Nanotubes Assemble from Amorphous Nanoparticles. *Journal of American Chemical Society*, 129(21):6820–6826, 2007.

[9] Wei-Qiang Han, Wen Wen, Ding Yi, Zhenxian Liu, Mathew M. Maye, Laura Lewis, Jonathan Hanson, and Oleg Gang. Fe-Doped Trititanate Nanotubes: Formation, Optical and Magnetic Properties, and Catalytic Applications. *Journal of Physical Chemistry C*, 111(39):14339–14342, 2007.

[10] P. P. Edwards, P. A. Anderson, L. J. Woodall, A. Porch, and A. R. Armstrong. Can We Synthesise a Dense Bundle of Quasi One-Dimensional Metallic Wires? *Materials Science & Engineering. A, Structural Materials : Properties, Microstructure and Processing*, 217-218:198–202, 1996.

[11] P. D. Cradwick, K. Wada, J. D. Russel, N. Yoshinaga, C. R. Masson, and V. C. Farmer. Imogolite, A Hydrated Aluminium Silicate of Tubular Structure. *Nature Physical Science*, 240:187–189, 1972.

[12] L. A. Bursill, J. L. Peng, and L. N. Bourgeois. Imogolite: An Aluminosilicate Nanotube Material. *Philosophical Magazine A*, 80(1):105–117, 2000.

[13] Suchitra Konduri, Sanjoy Mukhejee, and Sankar Nair. Controlling Nanotube Di-

mensions: Correlation between Composition, Diameter, and Internal Energy of Single-Walled Mixed Oxide Nanotubes. *ACS Nano*, 1(5):393–402, 2007.

[14] Luciana Guimaraes, Andrey N. Enyashin, Johannes Frenzel, Thomas Heine, Helio A. Duarte, and Gotthard Seifert. Imogolite Nanotubes: Stability, Electronic, and Mechanical Properties. *ACS Nano*, 1(4):362–368, 2007.

[15] Suchitra Konduri, Sanjoy Mukherjee, and Sankar Nair. Strain Energy Minimum and Vibrational Properties of Single-Walled Aluminosilicate Nanotubes. *Physical Review B*, 74(3):033401, 2006.

[16] R. Saito, G. Dresselhaus, and M. S. Dresselhaus. *Physical Properties of Carbon Nanotubes*. Imperial College Press, 1998.

[17] Fernardo Alvarez-Ramirez. Ab Initio Simulation of the Structural and Electronic Properties of Aluminosilicate and Aluminogermanate Natotubes with Imogolite-like Structure. *Physical Review B*, 76(12):125421, 2007.

[18] F. Cora C. R. A. Catlow and A. A. Sokol. Electron Spin Localization and Correlation Effects for Point Defects in Semi-ionic Solids. *Computational Materials Science*, 17(2-4):312–318, 2000.

[19] G. Kresse and J. Hafner. Ab Initio Molecular Dynamics for Liquid Metals. *Physical Review B*, 47(1):558–561, 1993.

[20] G. Kresse and J. Furthmuller. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Physical Review B*, 54(16):11169–11186, 1996.

[21] P. E. Blochl. Projector Augmented Wave Method. *Physical Review B*, 50(24):17953–17979, 1994.

[22] G. Kresse and D. Joubert. From Ultrasoft Pseudopotentials to the Projector Augmented Wave Method. *Physical Review B*, 59(3):1758–1775, 1999.

[23] John P. Perdew and Yue Wang. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. *Physical Review B*, 45(23):13244–13249, 1992.

[24] Hendrik J. Monkhorst and James D. Pack. Special Points for Brillouin Zone Integrations. *Physical Review B*, 13(12):5188–5192, 1976.

[25] Thomas R. Cundari and Walter J. Stevens. Effective Core Potential Methods for the Lanthanides. *Journal of Chemical Physics*, 98(7):5555, 1992.

[26] Axel D. Becke. Density Functional Thermochemistry III. The Role of Exact Exchange. *Journal of Physical Chemistry*, 98(7):5648–5652, 1993.

[27] M. J. Frisch et al. Gaussian 03, Revision C.02, 2004. Gaussian, Inc., Wallingford, CT, 2004.

[28] John Petter Gustafsson. The Surface Chemistry of Imogolite. *Clays and Clay Minerals*, 49(1):73–80, 2001.

[29] Hong Li, S. D. Mahanti, and Thomas J. Pinnavaia. Modeling the Local Structure and Energetics of Protozeolitic Nanoclusters in Hydrothermally Stable Aluminosilicate Mesostructures. *Journal of Physical Chemistry B*, 109(7):2679–2685, 2005.

[30] Georg Kresse, Michael Schmid, Evelyn Napetschnig, Maxim Shishkin, Lukas Kohler, and Peter Varga. Structure of the Ultrathin Aluminum Oxide Film on NiAl(110). *Science*, 308(5727):1440–1442, 2005.

[31] Karen Hemelsoet, David Lesthaeghe, Veronique Van Speybroeck, and Michel Waroquier. Global DFT-Based Reactivity Indicators: An Assessment of Theoretical Procedures in Zeolite Catalysis. *Journal of Physical Chemistry C*, 111(7):3028–3037, 2007.

[32] V. E. Henrich and P. A. Cox. *The Surface Science of Metal Oxides.* Cambridge University Press, 1996.

[33] P. A. Tipler and G. P. Mosca. *Physics for Scientists and Engineers.* W. H. Freeman & Co Ltd, 2007.

[34] Lev Kantorovich. *Quantum Theory of Solid State: An Introduction.* Kluwer Academic Publishers, 2004.

[35] G. Rodrigue. Vectorizing the FFTs in Parallel Computations. *Academic Press*, page 51, 1982.

[36] E. Hernandez, C. Goze, P. Bernier, and A. Rubio. Elastic Properties of C and BxCyNz Composite Nanotubes. *Physical Review Letters*, 80(20):4502–4505, 1998.

[37] E. Hernandez, C. Goze, P. Bernier, and A. Rubio. Elastic Properties of Single-Wall

Nanotubes. *Applied Physics A: Materials Science & Processing*, 68(3):287–292, 1999.

[38] Jian Ping Lu. Elastic Properties of Carbon Nanotubes and Nanoropes. *Physical Review Letters*, 79(7):1297, 1997-1300.

[39] P. Ravindran, R. Vidya, P. Vajeeston, A. Kjekshus, and H. Fjellvag. Ground- and Excited-State Properties of Inorganic Solids From Full-Potential Density-Functional Calculation. *Journal of Solid State Chemistry*, 176(2):338–374, 2003.

[40] Gilberto Teobaldi et al. . Hydroxyl Vacancies in Single-Walled Aluminosilicate and Aluminogermanate Nanotubes. *Journal of Physics: Condensed Matter*, 21(19):195301, 2009.

Chapter 5

The Surface Chemistry of Rutile $TiO_2(110)$

5.1 Introduction

Titanium dioxide is a material of significant technological importance due to its photocatalytic activity [1] and its wide range of applications in industry. Because the Rutile $TiO_2(110)$ surface is thermodynamically stable and easy to prepare in the laboratory, it has become increasingly popular, emerging as the model metal-oxide for surface science studies [2]. The Rutile $TiO_2(110)$ surface, see Fig. 5.1, consists of alternate rows of five-fold coordinated Ti atoms and two-fold coordinated bridging O atoms along the [001] direction [2]. Common point defects on the surface include missing bridging oxygen atoms, (O_b), known as oxygen vacancies (O_{vac}), oxygen ad-atoms (O_{ad}) above five-fold coordinated Ti atoms (Ti_{5c}), terminal hydroxyls (OH_t) i.e hydroxyls adsorbed on top of Ti_{5c}, and hydrogen ad-atoms on top of O_b referred to as bridging hydroxyls (OH_b). In STM images, the five-fold coordinated Ti_{5c} atoms appear as bright rows while O_{br} rows appear dark. Oxygen vacancies (O_{vac}) and surface hydroxyls (OH_{br}) appear as bright spots between the bright Ti_{5c} rows [2].

Under surface preparation, which involves Ar-ion sputtering and annealing to temperatures up to 1100 K in ultra high vacuum (UHV), oxygen vacancies are inevitably introduced to the surface. Due to residual water present in the vacuum chamber under experimental conditions, interaction of H_2O with the surface is always present and the need of understanding the exact mechanism of the reaction between water and surface

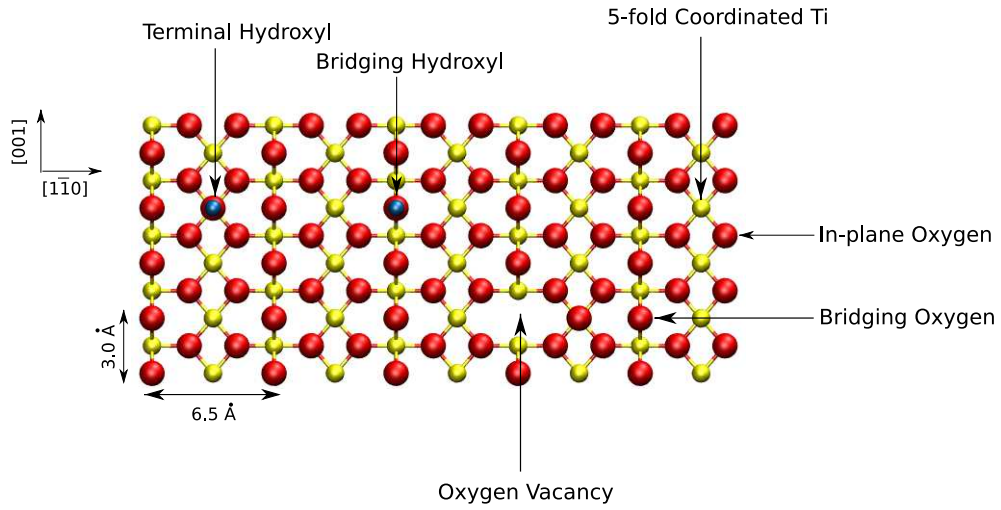

Figure 5.1: The Rutile $TiO_2(110)$ surface. O: red, H: blue, Ti: yellow.

oxygen vacancies has been increasingly pressing. Water is known to dissociate on oxygen vacancies by proton transfer to a nearby oxygen eventually creating two bridging hydroxyls [3]. It was later found that the ensuing change, after vacancy formation, in the oxidation state of the underlying Ti atoms from 4+ to 3+, leads to an inequivalence between the hydroxyl species after water dissociation [4]. Moreover, oxygen vacancies also facilitate the dissociation of molecular oxygen O_2 [5, 6], leaving a healed O_{vac} and an oxygen ad-atom O_{ad} on Ti_{5c} as described by the following reaction [7],

$$2Ti^{3+} + O_{vac} + O_2^0(g) \rightarrow 2Ti^{4+} + {O_b}^{2-} + {O_{ad}}^0. \tag{5.1.1}$$

The reduction of one oxygen atom of $O_2(g)$ in 5.1.1 to one bridging oxide species O_b^{2-} is accomplished by oxidation of the two Ti^{3+} sites associated with O_{vac} to Ti^{3+} [7]. The interaction between O_2 and bridging hydroxyls OH_{br}, however, is not very well understood yet and will be the main focus of the discussion in this chapter. The reaction path observed in experiments is

$$2Ti^{3+} + 2OH_{br}^{1-} + O_2^0(g) \rightarrow 2Ti^{4+} + H_2O^0(g) + {O_b}^{2-} + {O_{ad}}^0 \tag{5.1.2}$$

where the two Ti^{3+} species provide the two electrons necessary to reduce one oxygen atom of $O_2(g)$ to $H_2O(g)$ [7]. Theoretical calculations however do not predict water desorption at temperatures less than $350K$ [8]. In this chapter it will be discussed how negatively charged oxygen vacancies can open the possibility for alternative reaction channels and bridge the gap between the experimental evidence, obtained by tempera-

ture programmed desorption (TPD), of H_2O being desorbed at 310K [7, 9] and current ab initio studies favouring H_2O_2 as the most stable product of the reaction [8].

Before embarking in the theoretical analysis of the reactivity of the Rutile $TiO_2(110)$ surface, it should be acknowledged that the images, data and analysis appearing in this chapter relating to experiments performed in parallel with the theoretical simulations where provided by Prof. Geoff Thornton's group. They are mentioned here to motivate the discussion for the theoretical study following in the sections below. Assistance in the DFT calculations and STM simulations by Dr. Gilberto Teobaldi is gratefully acknowledged.

5.2 Experimental Results

Fig. 5.2B shows a surface containing both O_{vac} and OH_b alongside the same surface in Fig. 5.5C after it was exposed to 90 Langmuirs O_2 at 300 K. A number of small, bright spots can be seen on the Ti_{5c} sites (bright rows) in the latter image. The histogram of the height distribution of these bright spots, shown in Fig. 5.2D, indicates that these bright spots are almost entirely due to one final product. It should be noted here that at lower O_2 exposures a number of different types of species can be seen on Ti_{5c} rows that are likely to arise from terminal hydroxyls (OH_t) and other metastable species such as O_2H. These latter results are consistent with previous works [9, 10]. Analysis of the images in Fig. 5.2 shows that before the reaction with O_2 there were about 72 OH_b species and 38 O_{vac}s. Following the reaction with O_2, about 118 new bright spots can be seen on the bright Ti_{5c} rows in Fig. 5.2C. As it is well known that O_{ad} form after exposure of O_{vac} to O_2 [5, 11–14], it can be concluded that O_{ad} also forms from exposure of OH_b to O_2, in line with the conclusions of Henderson et al. [7]. Moreover, the availability of only 72 H atoms rules out the possibility that the products in the present case are OH_t [9, 10]. There can be noted, however, a subtle but crucial discrepancy between these observations and the conclusions of Henderson et al. [7]. With a starting point of 38 O_{vac} and 72 OH_b, it is expected that each O_{vac} yields one O_{ad} (reaction 5.1.1) and each OH_b yields half an O_{ad} (reaction 5.1.2), i.e. only 74 O_{ad} in total compared to the 118 found in the experiment. Since the availability of electrons from reduced TiO_2 is a prerequisite for O_2(g) dissociative adsorption on $TiO_2(110)$ [7, 10], the extra O_{ad} found in Fig. 5.2C requires the TiO_2 sample to supply more electrons than expected on the basis of the

number of O_{vac} ($2Ti^{3+}$) and OH_b ($1Ti^{3+}$) initially present. Thus, the assignment of the adsorbates in Fig. 5.2C to O_{ad} as well as the interpretation of TPD results [7] challenges the current theoretical understanding of the surface chemistry of TiO_2(110). Even by using a hybrid HSE06 approach [15] for this study, which is expected to describe more accurately than other functionals the electronic structure of metal oxides, H_2O_2 was predicted to be the overwhelmingly most stable product of the reaction between O_2 and OH_b. This result, which is inconsistent with the TPD data, mimics that found in previous simulations [10]. The electronic charge distribution associated with O_{vac} was

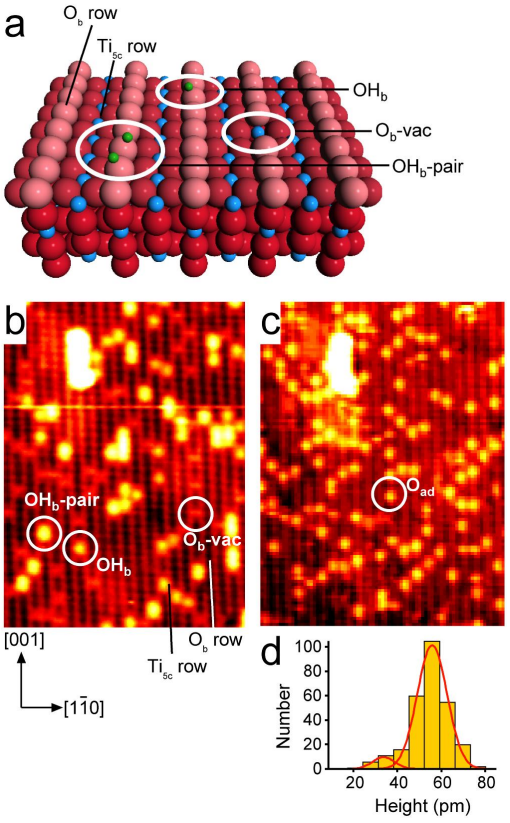

Figure 5.2: Reaction of O_2 with TiO_2(110) (A) Ball model of TiO_2(110). Red and blue spheres denote O and Ti, respectively. The pink spheres are bridging O atoms, which lie in the [001] azimuth of the substrate. Parallel Ti rows that lie between the bridging O rows are five-fold coordinated Ti atoms. Green spheres indicate H atoms (from OH_b). (B) $130 \times 170 Å^2$ STM image (V = 1.5 V, I = 0.25 nA) of an as-prepared TiO_2(110) surface which contains O_{vac} and OH_b. OH_b forms from dissociation of water from the residual vacuum at O_{vac}. An O_{vac}, an OH_b, and an OH_b pair are indicated. (C) The surface in B after exposure to 90 L O_2 at 300 K. One of the bright spots assigned to O_{ad} is circled. B-C have been smoothed using Image SXM [16] v.1.75. (D) A histogram showing the height distribution of 276 bright spots found on the Ti_{5c} rows fitted to two Gaussian curves. The data are taken from an unsmoothed, larger version of the image in C. The histogram indicates that the reaction products are almost entirely from one species.

also probed experimentally using scanning tunneling spectroscopy (STS) at 5 K. This low temperature was chosen to deliberately freeze out polaronic hopping effects which

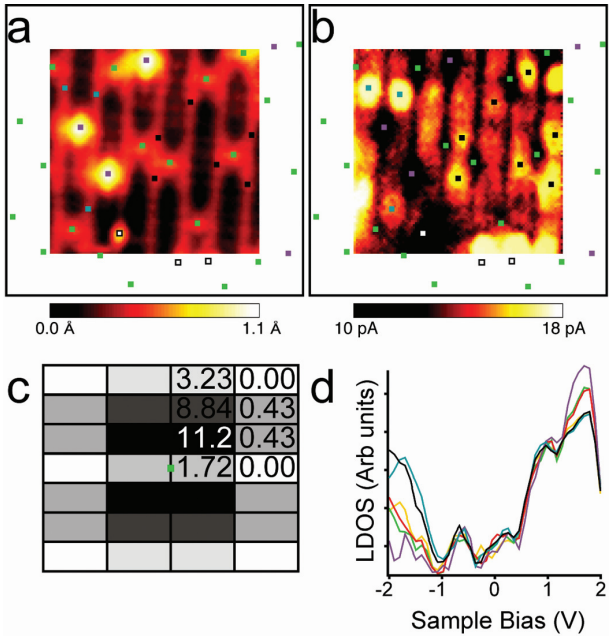

Figure 5.3: Experimental STM and STS data. (A) $(44\text{Å})^2$ STM image recorded simultaneously with the STS. (B) $(44\text{Å})^2$ CITS current map at -2 V. The squares in A and B show the positions of O_{vac} (green), OH_b (purple), and some bright features associated with O_{vac} (black). One impurity is also present and marked with a white square. Using a larger-scale image, the positions of O_{vac}, OH_b, and other impurities outside the area imaged in A and B are also indicated. A and B have been smoothed using Image SXM [16] v.1.75. (C) A correlation map between O_{vac} and bright features in B. The center of the map represents the position of an O_{vac} shown as a green square. The black rectangles represent unit cells centered on Ti_{5c} atoms that surround O_{vac}. The results are averaged between the four quadrants with the numbers shown only in one quadrant. The results are expressed as percentages which add to 100% when the numbers in all four quadrants are summed. The darker the shading, the greater the probability of finding a bright feature at the separation indicated by the map. (D) STS spectra represented as LDOS plots by plotting $(dI/dV) \times (V/I)$ vs V (26). The LDOS plots are taken from the bright features associated with O_{vac} (black), O_{vac} (green), OH_b (purple), Ti_{5c} (yellow), and O_b (red). Each curve is averaged from 180 individual spectra taken from the CITS set shown in this work and another 180 individual spectra taken from an equivalent CITS set recorded in an identical area of the surface. The black squares in A and B indicate which bright features contribute to the curves for bright features associated with O_{vac}; none were counted when they were also diagonally adjacent to OH_b or in close vicinity to impurities.

are thought to average out asymmetries that are predicted to arise at 0 K [17–19]. During the acquisition of the STM image shown in Fig. 5.5A, the tip was immobilised at each point of the scan while the tunneling currents versus bias voltage (I-V) spectra were recorded. This method of recording I-V spectra alongside STM images is known as current imaging tunneling spectroscopy (CITS) [20], and it allows the I-V spectra to be displayed as current maps at each voltage. The STM image in Fig. 5.5A was acquired at a sample bias of +2 V with a tunneling current of 0.03 nA. It clearly shows both OH_b and O_{vac} as bright spots between bright Ti_{5c} rows. The current map at +2 V has a similar appearance to the STM image, as one would expect. However, the current map at -2 V (Fig. 5.5B) has a very different appearance. O_{vac} and OH_b both appear as dark spots centred between two bright rows. Bright features can be seen on the bright Ti_{5c} rows and a correlation analysis (Fig. 5.5C) shows that these lie diagonally adjacent to the O_{vac} in the positions of the second nearest Ti_{5c} neighbours. Each O_{vac} tends to be surrounded by one or two brighter features rather than four equally bright features. Fig. 5.5D shows normalised conductance spectra $(dI/dV)\times(V/I)$ which approximate the local density of states. Each curve represents an average of 360 spectra corresponding to the bright features, O_{vac}, OH_b, Ti_{5c}, and O_b. Similar spectra were recorded using the same tip with the tunneling current set between 0.015 nA and 0.7 nA. Some spectra were also recorded using an iridium tip. In neither case did however was observed a shift of features, which suggests that the effect produced by tip-induced band-bending is negligible. These STS data are in broad agreement with recently reported 78 K results [21].

5.3 Methodology

Density Functional Theory (DFT) calculations were performed using the Vienna Ab-Initio Simulation Package (VASP) [22, 23] using a plane-wave basis set with an energy cut-off of 450 eV. The atom cores were described using the Projector Augmented Wave method (PAW) [24, 25] were for the Ti atoms the $3s$ and $3p$ states were treated as valence states. Slabs of two different sizes were used during the calculations with different sampling grids for the Brillouin zone in each case. The 3×2 slabs were constructed using four layers of $O - TiO_2 - O$ with the atoms in the upper two layers free to move and those in the bottom two layers fixed in their bulk positions. The sampling of the

Brillouin zone was performed using a $4 \times 3 \times 1$ Monkhorst-Pack grid of special k-points. The values for the 2×2 slabs were five layers with the bottom three fixed and a $6 \times 3 \times 1$ grid for the BZ sampling. Atomic relaxations were performed until all forces components of all atoms were less than 0.01 eV/Å. The distance between the slabs and their replicated images along the z-axis was 15Å which was achieved by introducing a 15Å thick layer of vacuum above them. The generalized gradient approximation (GGA) as implemented in the PW91 exchange-correlation functional [26] was adopted for the pure DFT calculations. The relaxed geometries were then used for single-point calculations using the hybrid HSE03 functional due to Heyd, Scuseria and Ernzerhof, in order to produce accurate wavefunctions before performing to the STM and STS simulations. Due to the increased computational costs of performing hybrid DFT calculations with a plane-wave basis set, full relaxations with this functional were not possible. Scanning Tunneling Microscopy simulations were performed using the code bSKAN [27, 28] within both the Tersoff-Hamann [29, 30] and Bardeen [31] approximations. LSDA+U calculations where performed using a local $U_{\text{eff}} = 5.5$ eV correction for Ti atoms as suggested by embedded cluster configuration interaction results [32].

5.4 Defect Formation Energies

Defect formation energies were calculated using the scheme presented in [33]. In this scheme, the formation energy of a charged defect, within a formulation were the surface is allowed to exchange energy and particles with a reservoir, is given by

$$E_f(X^q) = E_X(q) - E_{\text{clean}} - \sum_i n_i \mu_i - q(E_f - E_{\text{VBM}} + \Delta V) + \Delta F_{\text{vib}}. \tag{5.4.1}$$

$E_X(q)$ is the total energy as calculated by first-principles DFT calculations for a defect X with charge q, while E_{clean} the total energy for the pristine surface. Obviously, for neutral defects the fourth term in Eq. 5.4.1 vanishes as $q = 0$. $\Delta\mu_e = E_f - E_{\text{VBM}} + \Delta V$ is the change in the electronic chemical potential induced by the exchange of electrons between the charged defect and the bulk crystal. The number of atoms with chemical potential μ_i that had to be extracted or added to the surface, in order to create the defect, are given by n_i. The chemical potential μ_i corresponds to the energy needed when the atom is exchanged with the reservoir in order to create the defect. The term $E_f - E_{\text{VBM}}$ refers to the system's Fermi energy referenced to the valence band maximum.

The terms ΔV and ΔF_{vib} refer to the shift of the electrostatic potential and the change in the vibrational free energy respectively after introducing the defect. By neglecting temperature broadening effects, the electronic chemical potential μ_e corresponds to the Fermi energy level E_f and therefore the term $E_f - E_{\text{VBM}} + \Delta V$ in Eq. 5.4.1 is the change induced in the electronic chemical potential by the shift ΔV in the electrostatic potential under the presence of the charged defect X^q. The change of the vibrational free energy in particular is given by

$$\Delta F_{\text{vib}} = \sum_i^{N_X} \Big[\frac{1}{2}hcu_i^X + kT\ln\Big(1 - e^{-\frac{hcu_i^X}{kT}}\Big)\Big] - \sum_i^{N_{\text{clean}}} \Big[\frac{1}{2}hcu_i + kT\ln\Big(1 - e^{-\frac{hcu_i}{kT}}\Big)\Big] \quad (5.4.2)$$

where N_X and N_{clean} are the modes of wavenumber (cm^{-1}) u^X and u for the surfaces with and without the defect respectively. It can be seen in Eq. 5.4.2 that the terms of form $\frac{1}{2}hcu$ inside the brackets consist of the system's vibrational zero point energy while the other two terms of the contributions of the vibrations to the surface's entropy. Returning to Eq. 5.4.1 and the chemical potential μ_i, it can be argued that the following relation between μ_i and the partial pressure in the vacuum chamber can be established

$$\mu_O = \frac{1}{2}E(O_2) + \frac{1}{2}kT\Big\{\ln\Big[\frac{P_{O_2}V_{O_2}}{kT}\Big] - \ln Z_{rot} - \ln Z_{vib}\Big\} \quad (5.4.3)$$

with V_{O_2} the quantum volume

$$V_{O_2} = \Big[\frac{\hbar}{m_{O_2}kT}\Big] \quad (5.4.4)$$

for O_2. Similar expressions with Eq. 5.4.3 and 5.4.4 hold for H_2O and H_2. Z_{rot} and Z_{vib} are the rotational and vibrational partition functions and P_{O_2} the oxygen partial pressure. E_{O_2} is the energy of a single O_2 molecule. Under thermodynamic equilibrium, $\mu_{H_2O} = 2\mu_H + \mu_O$ and therefore

$$\mu_H = \frac{1}{2}\Big[\mu_{H_2O} - \mu_O\Big] \quad (5.4.5)$$

which allows for the evaluation of the chemical potential as a function of the temperature T and the partial pressures P_{O_2} and P_{H_2O}. The terms in Eq. 5.4.2 can be rearranged by separating the contributions to the vibrational free energy from the surface and the adsorbate that introduces the defect. By labeling the vibrations due to the adsorbate

as N_{ads} and the surface as $N_{\text{surf}} = N_X - N_{\text{ads}}$ Eq. 5.4.2 receives the following form

$$\begin{aligned}\Delta F_{\text{vib}} = &\sum_{i}^{N_{\text{ads}}} \left\{ \frac{1}{2} hcu_i^X + kT \ln \left[1 - \exp\left(-\frac{hcv_i^X}{kT} \right) \right] \right\} \\ &+ \sum_{i}^{N_X - N_{\text{ads}}} \left\{ \frac{1}{2} hcu_i^X + kT \ln \left[1 - \exp\left(-\frac{hcv_i^X}{kT} \right) \right] \right\} \\ &- \sum_{i}^{N_{\text{clean}}} \left\{ \frac{1}{2} hcu_i^X + kT \ln \left[1 - \exp\left(-\frac{hcv_i}{kT} \right) \right] \right\} \end{aligned} \tag{5.4.6}$$

which in turn can be rewritten as

$$\begin{aligned}\Delta F_{\text{vib}} = &\sum_{i}^{N_{\text{ads}}} \left\{ \frac{1}{2} hcu_i^X + kT \ln \left[1 - \exp\left(-\frac{hcv_i^X}{kT} \right) \right] \right\} \\ &+ \sum_{i}^{N_{\text{clean}}} \left\{ \frac{1}{2} hc(u_i^X - u_i) + kT \ln \left[\frac{1 - \exp\left(-\frac{hcv_i^X}{kT} \right)}{1 - \exp\left(-\frac{hcv_i}{kT} \right)} \right] \right\} \\ &= \Delta F_{\text{ads}} + \Delta F_{\text{pert}}. \end{aligned} \tag{5.4.7}$$

The first term in Eq. 5.4.7 represents the contribution to the vibrational free energy ΔF_{vib} stemming from the vibrations induced on the surface by the defect X, the second term is due to vibrational modes induced on the substrate. According to studies that have been performed on the subject of NH_3 interacting with GaN surfaces, the contribution of ΔF_{pert} in Eq. 5.4.7 was found to not contribute substantially in the formation energy, with values of less than 0.1eV [33–35]. Eq. 5.4.2 can eventually then be written as

$$\begin{aligned}E_f(X^q) = &E_X(q) - E_{\text{clean}} - \sum_i n_i \mu_i + q(E_f - E_{\text{VBM}} + \Delta V) + \\ &+ \sum_{i}^{N_{\text{ads}}} [\frac{1}{2} hcu_i^X + kT \ln(1 - e^{-\frac{hcv_i^X}{kT}})] \end{aligned} \tag{5.4.8}$$

in the case where surface defects are overally charged. In the case where the simulated system is neutral, Eq. 5.4.8 reduces to

$$E_f(X^0) = E_X(0) - E_{\text{clean}} - \sum_i n_i \mu_i + \sum_{i}^{N_{\text{ads}}} \Big[\frac{1}{2} hcu_i^X + kT \ln\Big(1 - e^{-\frac{hcv_i^X}{kT}}\Big) \Big]. \tag{5.4.9}$$

Returning to Eq. 5.4.1, it can be noticed that the dependence of the formation energy to the charge q accumulated in the defect, is taken into account through the term $-q\Delta\mu_e = -q(E_f - E_{\text{VBM}} + \Delta V)$. The term $-q\Delta\mu_e$ corresponds to the electron exchange

with the Fermi energy of the bulk. Therefore the Fermi energy of the system with the charged defect has to be referenced to the valence band maximum, E_{VBM}, of the bulk rutile TiO_2. In order for the change in the chemical potential μ_e to be evaluated appropriately, the electrostatic potential of the charged slab needs to be aligned to the unpolarized, neutral bulk reference. The potential shift induced by the presence of the charged defect can be evaluated using the following technique. The shift in the electrostatic potential ΔV can be evaluated by referring to the transversely averaged dipole surface density μ_σ

$$\Delta V = 4\pi\mu_\sigma \tag{5.4.10}$$

using the following expression

$$\mu_\sigma = \int_{z_0-c/2}^{z_0+c/2} dz(z - z_0)\bar{\rho}(z) \tag{5.4.11}$$

were z_0 is the system centre of charge, c is the dimension of the simulation cell along the axis perpendicular to the surface of the 2D-periodic slab. The transversely averaged global charge density $\bar{\rho}$ is given by

$$\bar{\rho}(z) = \frac{1}{S}\int_S dxdy\rho(\mathbf{r}) \tag{5.4.12}$$

with S being the surface area of the simulated slab on the xy plane. $\bar{\rho}(z)$ can be obtained by solving Poisson's equation

$$\frac{d^2\bar{V}(z)}{dz^2} = -4\pi\bar{\rho}(z) \tag{5.4.13}$$

with

$$\bar{V}(z) = \frac{1}{S}\int_S dxdyV(\mathbf{r}). \tag{5.4.14}$$

From a practical point of view, the Fast Fourier Transformation (FFT) coefficients of the transversely averaged potential

$$\bar{V}_G = \sum_{k=0}^{N} \bar{V}_k e^{i2\pi Gk/N} \tag{5.4.15}$$

are calculated on the same grid N as with the DFT simulation. The FFT coefficients of the transversely averaged charge density were obtained from

$$\bar{\rho}_0 = 0 \tag{5.4.16}$$

$$\bar{\rho}_G = \frac{G^2}{4\pi}\bar{V}_G \tag{5.4.17}$$

and then $\bar{\rho}$ was evaluated by FFT back transformation. Calculated values for the defects considered can be seen in Table 5.1. The calculated transversely averaged charge density

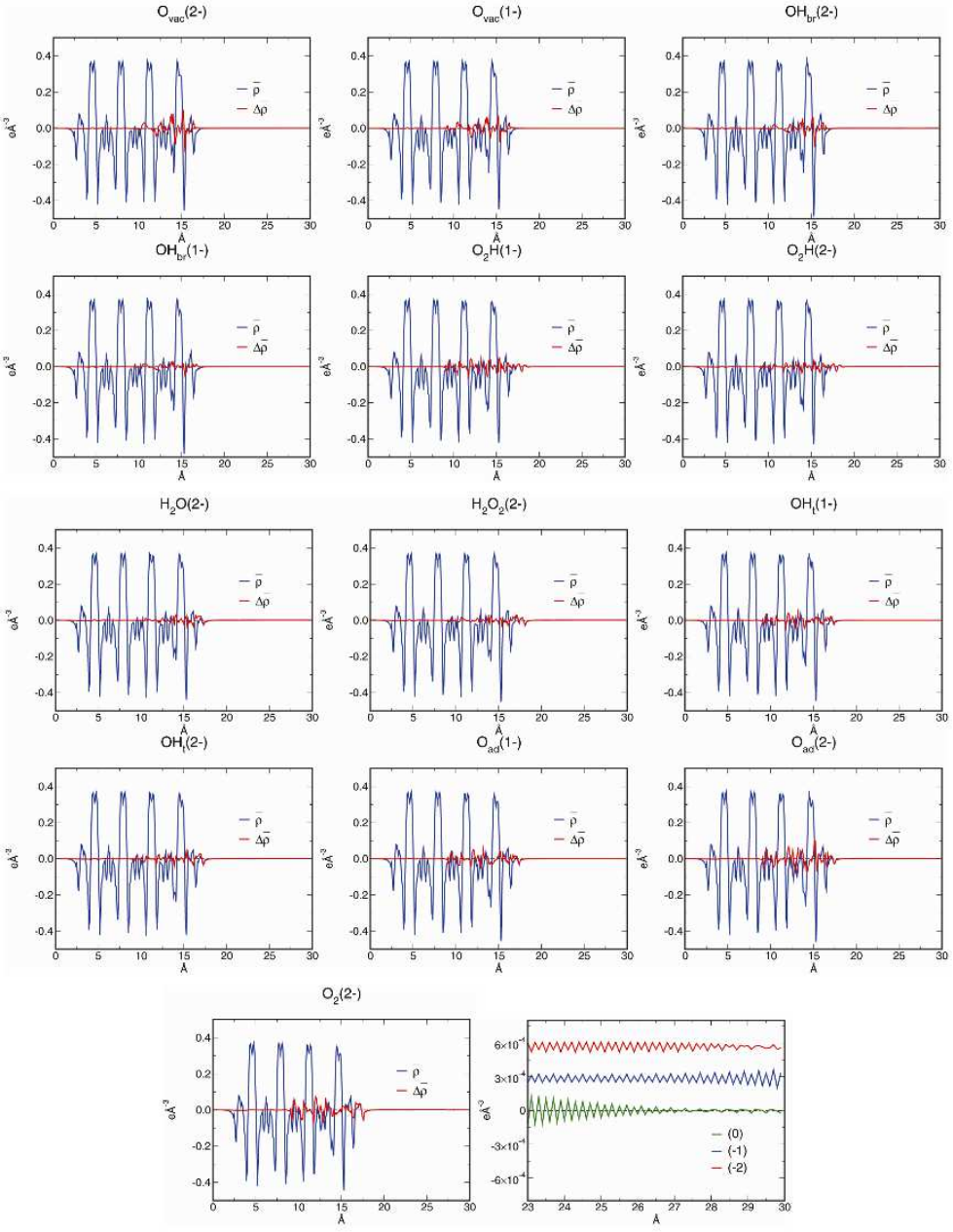

Figure 5.4: Transversely averaged charge density $\bar{\rho}$ and corresponding change with respect to a clean $TiO_2(110)$ slab for all considered charged systems. Simulation cell volume: $13.1692 \times 8.91 \times 30$Å.

$\bar{\rho}$ for all the considered charged systems can be seen in Fig. 5.4, along with the change $\Delta\bar{\rho} = \bar{\rho}_X - \bar{\rho}_{\text{clean}}$ between the considered slab and the clean surface. In the $6 - 8$Å range, the changes in the charge density are negligible proving that the thicknesses of the modeled slabs are adequate.

Vibrational frequencies were calculated using a scheme similar to the one by Northrup [36]. The total energy V_0 of the systems with the atoms at their equilibrium positions was initially calculated using the method described in the Methodology section. With the lattice potential energy V given by

$$V = V_0 + \frac{1}{2}\sum_{i,j} u_i K_{ij} u_j \tag{5.4.18}$$

where u_i is the displacement of the atoms from their equilibrium positions and K_{ij} the force constant matrix. In order to evaluate K_{ij}, atoms are displaced by a small amount in every spatial degree of freedom, usually along the x,y and z axes, and the forces induced are calculated using the Feynman-Hellmann theorem [37]. To include anharmonic effects, displacements are enforced on both directions of each degree of freedom and then the averaged value is used. By dividing the force induced on the i^{th} atom by the displacement of the j^{th} atom the force constant matrix K_{ij} was obtained. The dynamical matrix is then

$$D_{ij} = \frac{K_{ij}}{M_i M_j} \tag{5.4.19}$$

with the vibrational frequencies given after diagonalization of Eq. 5.4.19. The values calculated using the presented formalism for the considered defects are given in the table Tab. 5.1. The substantial contribution to ΔF_{vib} for the surfaces containing O-H bonds can be attributed to the zero point energy contribution of the high energy vibrations of these bonds like stretching and bending.

5.5 Electronic Structure Analysis

O_{ad} can be modeled as the most stable product only when excess electronic charge is allowed to accumulate at the O_{vac} and OH_b (Fig. 5.5), the latter being formed by reaction of water molecules with O_{vac}. The presence of extra electrons at O_{vac}/OH_b originates from occupation of additional 3d electron states arising from polaronic distortion at both O_{vac} and OH_b sites. As it will be discussed later, the excess electronic charge is

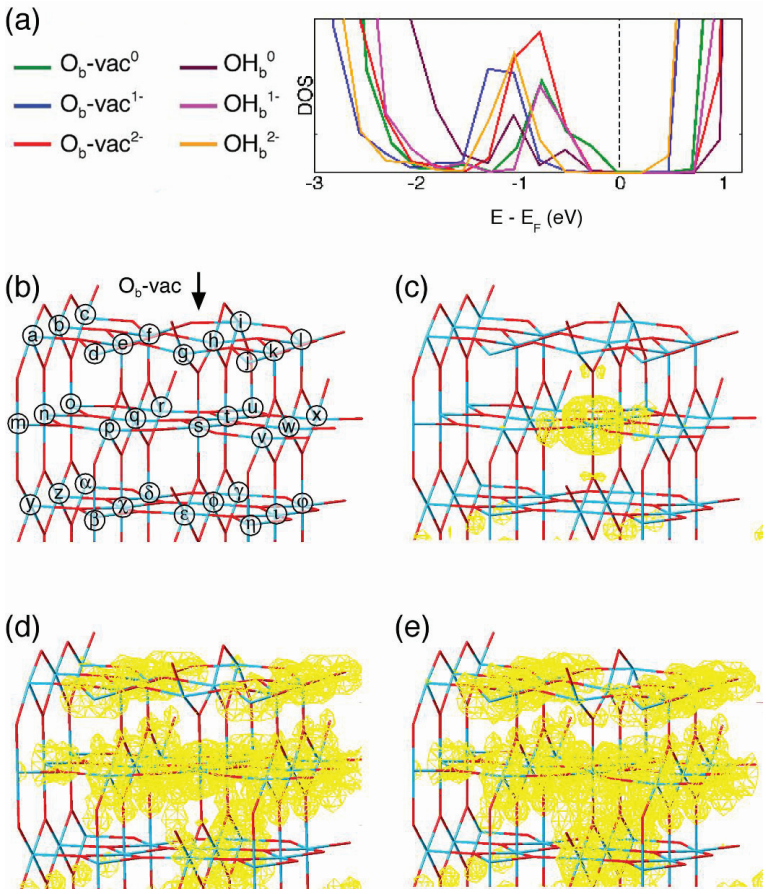

Figure 5.5: Calculated electronic structure of O_{vac} and OH_b. (A) The total density of states for the optimized layers in the presence of different amounts of extra electronic charge. (B) The $TiO_2(110)$ surface shown as a stick model where the blue intersections indicate Ti sites and the red intersections indicate O sites. The arrow points at the O_{vac} and each Ti atom is labeled. The global charge density of the BGS is shown in yellow ($10^{-6} e\text{Å}^{-3}$) for $O_{vac}(0)$, $O_{vac}(1-)$, and $O_{vac}(2-)$ in C, D, and E, respectively.

Table 5.1: Structural and energy parameters for the considered surfaces and defects. Rutile TiO_2(110) bulk $E_{VBM} = 2.26$eV. HSE03 functional.

X(q)	d(Å)[a]	ΔF_{vib}(eV)	E_{Fermi}(eV)	$\Delta\bar{V}$(eV)	$\Delta\mu_e$(eV)	E_f(eV)
O_{vac}^{-2}	1.94	-0.12	+0.08	+0.59	-0.90	+0.41
O_{vac}^{-1}	1.95	-0.12	-0.44	+1.80	-1.59	+2.75
O_{vac}^{0}	1.96	-0.12	-	-	-	+4.43
OH_{br}^{-2}	0.97	+0.31	+0.22	+1.12	-0.92	+0.15
OH_{br}^{-1}	0.97	+0.31	-0.76	+1.62	-1.40	+0.40
OH_{br}^{0}	0.97	+0.31	-	-	-	-1.89
$H_2O_2^{-2}$	2.36	+0.76	+0.09	+0.60	-1.57	-1.02
$H_2O_2^{0}$	2.27	+0.76	-	-	-	-1.89
O_2H^{-2}	1.88	+0.42	-0.51	+0.17	-2.60	-4.60
O_2H^{-1}	1.86	+0.42	-1.85	-0.62	-4.73	-3.71
O_2H^{0}	2.26	+0.42	-	-	-	+2.71
H_2O^{-2}	2.32	+0.63	+0.31	+1.02	-0.93	-0.83
H_2O^{0}	2.24	+0.63	-	-	-	+1.68
OH_{t}^{-2}	1.83	+0.31	-0.20	-0.59	-3.05	-6.75
OH_{t}^{-1}	1.81	+0.31	-2.35	-1.14	-5.75	-5.94
OH_{t}^{0}	1.79	+0.31	-	-	-	+2.45
O_{ad}^{-2}	1.64	+0.06	-1.63	-3.49	-7.58	-16.06
O_{ad}^{-1}	1.66	+0.06	-2.04	-3.69	-7.79	-6.84
O_{ad}^{0}	1.67	+0.05	-	-	-	-
O_2^{-2}	1.85	+0.05	-0.45	-3.37	-6.35	-13.01
O_2^{0}	-	-	-	-	-	+2.54

not pinned directly at the O_{vac} but is rather found localised at a number of reduced Ti sites around the vacancy (see also Fig. 5.6). The additional electronic charge generates a surface polaron. High affinities for both surface and bulk electron polaron trapping have been reported for substrates such as HfO_2 (12) and SiO_2 (13). This has been used to explain a number of surface properties, such as SiO_2 discharge phenomena, although not the surface chemistry. Various spectroscopies suggest that O_{vac} introduces electronic states into the band gap about 1 eV below the conduction band (CB) onset (3, 14, 15). These band gap states (BGS) are known to have d character (14). The BGS persist even when the O_{vac} are replaced with OH_b formed by water dissociation at the O_{vac} (3, 14). In these calculations, regardless of the number of electrons occupying the BGS [2 electrons (e) for $O_{vac}(0)$, 3 e for $O_{vac}(1-)$, 4 e for $O_{vac}(2-)$, 1 e for $OH_b(0)$, 2 e for $OH_b(1-)$, 3 e for $OH_b(2-)$] it was found that the energy remains close to 1 eV below the CB onset (Fig. 5.5A). The extra electrons, however, drastically affect the spatial distribution of the associated 3d electronic density. While for $O_{vac}(0)$, the simulations suggest a sub-surface localization immediately beneath the defect site (Ti_s in Fig. 5.5B), the presence of one and two extra electrons induces an electronic reorganization that moves electronic charge away from the O_{vac} site and the corresponding O_b row (as it

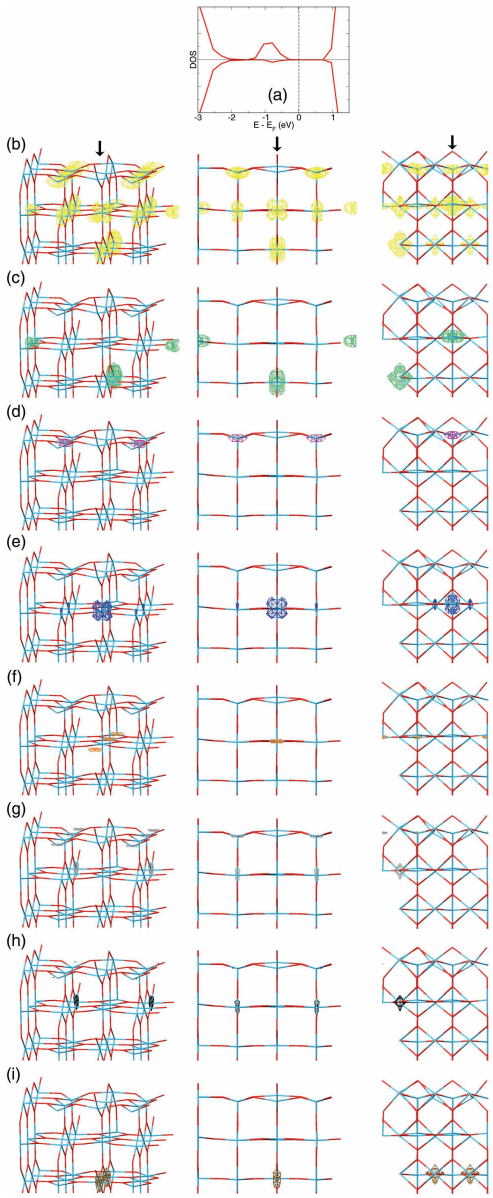

Figure 5.6: (a) Total Density of States (DOS) integrated only over relaxed atoms (red) for $O_{vac}(2-)$. (b) Electronic charge distribution (yellow: $10^{-5}e\text{Å}^{-3}$) summed up for all BGS shown from different perspectives: tilted (left), -oriented (middle) and -oriented (right). The integration has been performed over the whole Brillouin zone by considering the energy window spanned by the BGS peak in (a). (c)-(i) Single state-resolved electronic distribution ($10^{-5}e\text{Å}^{-3}$) for the BGS integrated as in (b). Blue intersections indicate Ti atoms and red intersections O atoms. The position of O_{vac} is marked by the black arrow.

can be seen in Fig. 5.5D-E). In line with other hybrid DFT results for O_{vac} in high-k

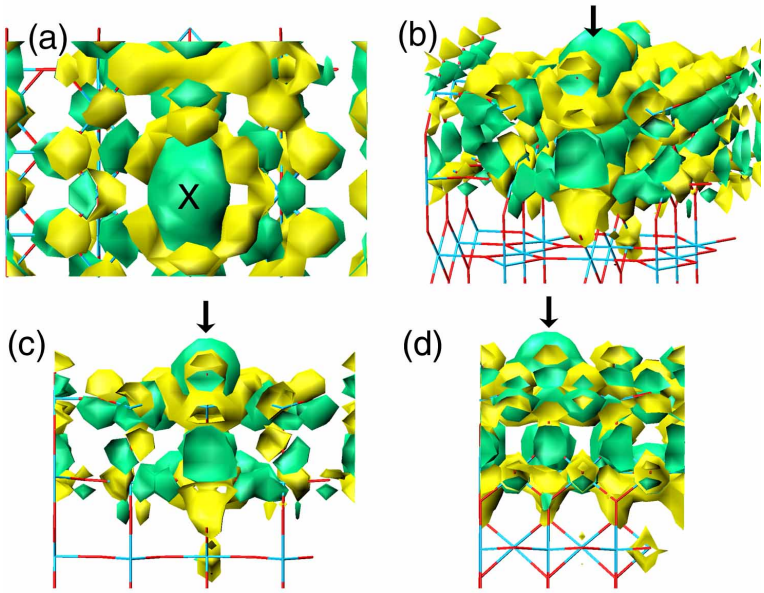

Figure 5.7: Asymmetric $O_{vac}(2-)$ induced global electron charge density accumulation ($\Delta\rho > 0$, yellow) and depletion ($\Delta\rho < 0$, green) regions, $\Delta\rho = \rho_X - \rho_{clean}$ [X = $O_{vac}(2-)$], viewed from different perspectives: (a) top, (b) tilted, (c) -oriented, (d) -oriented (right). All figures report an iso-contour of $10^{-5} e\text{Å}^{-3}$. The blue intersections indicate Ti atoms and the red intersections represent O atoms. The position of O_{vac} is indicated by the X mark in (a) and by the black arrow in (b)-(d).

dielectrics (12, 16), an analysis of the O_{vac} induced changes in the global electronic distribution reveals symmetry breaking at the defect site (details are shown in Fig. 5.8). This is discussed in more detail in the next section.

5.6 Ti Interstitials as Charge Donors

The possible origin of the additional electronic charge will be considered in this section. Given the ease of electron transfer in TiO_2 [18, 38] with reported barriers as low as 0.09 eV and mobilities as high as $5.24 \times 10^{-2} cm^2 V^{-2} s^{-1}$, it is reasonable to expect interstitial Ti (Ti_{int})-donated electrons to diffuse comparatively freely within the system [39]. To test the hypothesis that they will, in this case, diffuse towards the O_{vac}/OH_b sites at the surface, a much larger unit cell (8 tri-layers) was also modeled with inclusion of one initially neutral Ti_{int} atom. The increased size of the simulated systems here (288 atoms for the clean slab) prevented the application of HSE06 so a local spin density approximation (LSDA)+U scheme was adopted instead, already successfully applied to O_{vac} on TiO_2(110) [32]. Fig. 5.9 shows the BGS spatial distribution for O_{vac}, both isolated and in the presence of one Ti_{int}. Recently, Ti_{int} species have been suggested as

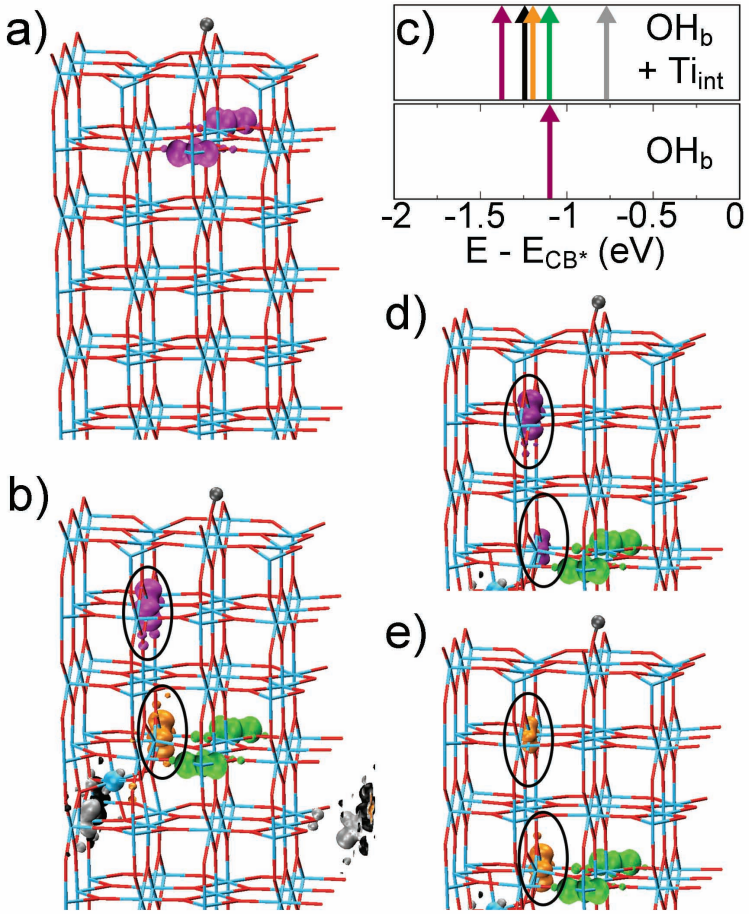

Figure 5.8: Electronic charge density distribution ($10^{-5} e\text{Å}^{-3}$) for the BGS of OH_b (a) and OH_b modeled together with one Ti_{int} between the 4th and 5th TiO_2(110) tri-layers ($OH_b + Ti_{int}$) (b). Blue intersections indicate Ti atoms and red intersections O atoms. Ti_{int} and H atoms are shown as blue and gray spheres, respectively. Different colours have been used to distinguish between Ti_{int}-donated (black, gray, green, orange), and surface OH_b-induced (purple) BGS. (c) Single-state energy level diagram with respect to the CB onset ($E - E_{CB*} = 0$) for the BGS in (a)-(b) after vacuum level electrostatic alignment. Up and down arrows refer to the modeled spin of the specific state. The same BGS colour labeling has been used for all the displayed (a)-(e) panels. For clarity, the hybridized BGS of OH_b+Ti_{int} (orange and purple, circled in (b) are re-displayed in (d)-(e) with the omission of the orange (d) and purple (e) BGS.

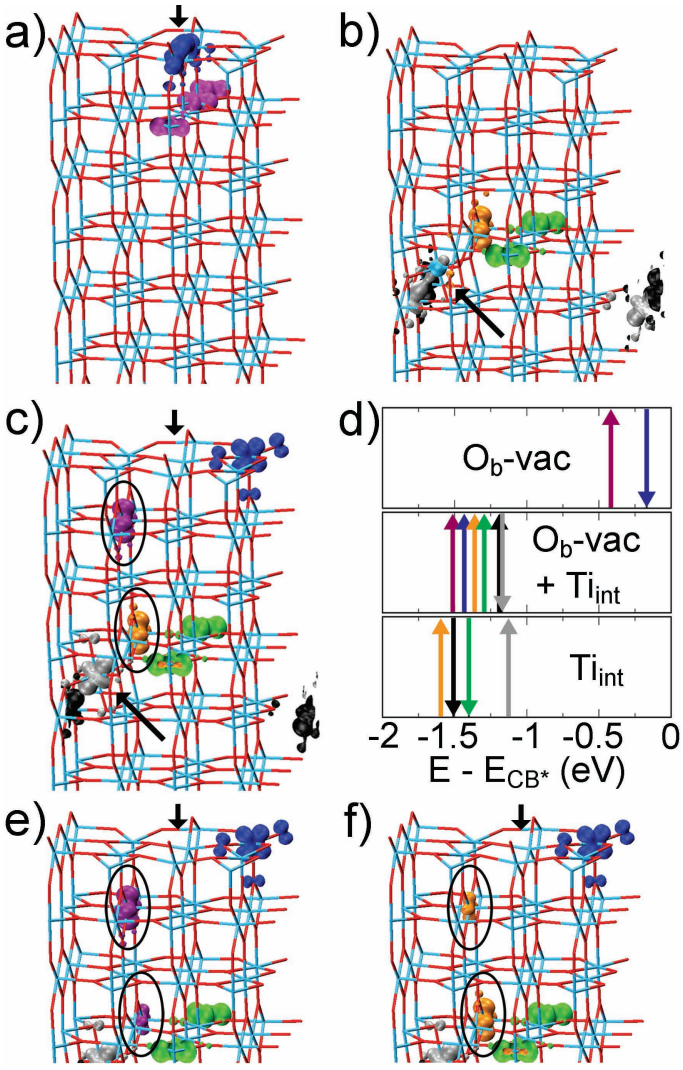

Figure 5.9: Electronic charge density distribution ($10^{-5}e\text{Å}^{-3}$) for the BGS of the considered 8 tri-layer systems. (A) O_{vac}, (B) one Ti_{int} between the 4th and 5th tri-layers, (C) O_{vac} modeled together with one Ti_{int} between the 4th and 5th $TiO_2(110)$ tri-layers ($O_{vac} + Ti_{int}$). Blue intersections indicate Ti atoms and red intersections O atoms. Ti_{int} is shown as a blue sphere. The positions of O_{vac} and Ti_{int} are marked by the black arrows. Different colors have been used to distinguish between Ti_{int}-donated (black, gray, green, orange), and surface O_{vac} induced (purple, blue) BGS. (D) Single-state energy level diagram with respect to the CB onset ($E - E_{CB*} = 0$) for the BGS in A-C after vacuum level electrostatic alignment. Up and down arrows refer to the modeled spin of the specific state. The same BGS color labeling has been used for all the displayed A-F panels. (E-F) For clarity, the hybridized BGS of O_{vac}+Ti_{int} (orange and purple) circled in C, are re-displayed with the omission of the orange in E and purple in F BGS.

the main origin of Titania BGS [40] with negligible importance of O_{vac} (or OH_b) sites. In contrast, the results shown in Fig. 5.9C-D demonstrate that a combination of Ti_{int} and O_{vac} induces new BGS. For isolated O_{vac}, the BGS (and the associated Ti^{3+} sites) are symmetrically localised at or immediately beneath the defect site (Fig. 5.9A). For one isolated Ti_{int} (Fig. 5.9B), the calculations suggest an asymmetric localisation of the BGS at the Ti_{int} site (one state) and on other separate, regular lattice Ti sites across the slab. Following refs. [12, 41], the appearance of localised BGS, and the ensuing local re-hybridization is assigned to the chemical reduction of specific Ti sites from their stoichiometric oxidation state (i.e. reduction from Ti^{4+} to Ti^{3+} in an oversimplified ionic model).

Upon consideration of an O_{vac} together with one Ti_{int}, coupling between the vacancy site and the Ti_{int}-donated electrons is evidenced by the large changes in the BGS eigenvalues (Fig. 5.9D). By comparing the formation energies of an isolated O_{vac} (+5.11 eV) and a Ti_{int}-coupled vacancy (+2.72 eV), the simulations suggest a net stabilization of 2.39 eV for O_{vac} within the simulation cell. The concentration of Ti_{int} modeled is about an order of magnitude higher than experimental estimates from the literature [39]. This allows to capture the qualitative behaviour but it makes addressing quantitatively the number of electrons globally transferred to the surface not possible. Given that the Ti_{int} concentration simulated is about an order of magnitude greater than experimental estimates [39], this coupling should be even more favourable in reality as the electrostatic repulsion between BGS is expected to be much lower. This BGS coupling drastically affects both the energy and, as was the case for the HSE06 modeling for $O_{vac}(2-)$ (Fig. 5.5), the spatial distribution of the BGS of O_{vac}. The presence of Ti_{int}-donated BGS induces an electronic reorganization that moves electronic charge away from the vacancy site and the corresponding O_b row (Fig. 5.9C). It should be noted that this is a different role for Ti_{int} than that claimed in an earlier study, where electronic states associated with Ti_{int} were responsible for the BGS [12]. Analysis of the individual BGS for the coupled systems provides insight into how the Ti_{int}-donated electrons couple to O_{vac} sites. As shown in Fig. 5.9C-F, when Ti_{int}-donated electrons are modeled together with O_{vac}, hybridization between specific BGS (circled in Fig. 5.9C,E,F) make Ti_{int}-donated electrons available at the O_{vac} site thus locally charging O_{vac} with the extra electrons necessary to account for the additional O_{ad} imaged in Fig. 5.2C. The same mechanism is also found for the interaction of Ti_{int}-donated electrons with OH_b, In this

case, comparison between the formation energies of OH_b (+0.32 eV) and OH_b/Ti_{int} (-1.05 eV) suggests a net stabilization of 1.37 eV. As with the O_{vac}/Ti_{int} system, the Ti_{int} electronic surplus is also characterized by symmetry breaking and hybridization between specific BGS (circled in Fig. 5.8). Although the number of electrons globally transferred to the surface is not possible to be addressed, estimates can be made based on the upper limit of reported Ti_{int} concentration; $2 \times 10^{19}cm^{-3}$, i.e. one per 800 TiO_2 unit cells [39]. In this estimation, typical O_{vac}/OH_b defect concentrations are assumed, of $5 \times 10^{13}cm^{-2}$ sample and we consider that each Ti_{int} generates four Ti^{3+} sites (Fig. 5.9B and [42]). This means that to singly (doubly) charge O_{vac}/OH_b one would only require the participation of 0.0017% (0.0034%) of the global number of Ti_{int}-donated electrons.

5.7 Scanning Tunneling Microscopy Simulations

Fig. 5.10 shows the performed STM and current map simulations of O_{vac} at +2 V and -2 V, respectively. These are derived from the HSE06 density of states. The simulations at +2 V are similar to each other, the highest current being found at the O_{vac} (Fig. 5.10A,C,E). In contrast, the simulations at -2 V are strongly dependent on the O_{vac} charge. For both $O_{vac}(0)$ and $O_{vac}(1-)$, the highest current is found in a spot (or spots) directly in line with the bridging O row, inconsistent with the experimental contrast. The experimental distribution of the defect states is only reproduced in simulations of $O_{vac}(2-)$ where the highest current is found on a second nearest Ti_{5c} neighbour with little current along the O_b row. The current map for $O_{vac}(2-)$ at -2 V was deconstructed by separating the contributions into slices with energy windows of 0.2 eV (Fig. 5.11). In this way, it is found that the asymmetry originates between energies of 1.1 and 1.5 eV, with the majority signal between 1.1 and 1.3 eV. This identifies the asymmetry in the current map with the BGS introduced by $O_{vac}(2-)$. This asymmetry is not observed in data recorded at 78 K (25). Presumably this arises because of significant polaronic hopping of the type recently identified at room temperature. STM and current map simulations based on the LSDA + U treatment of Ti_{int}/O_{vac} introduced above (Fig. 5.9C) also fit the experimental results well. The asymmetry can again be traced to the BGS of O_{vac}, in this case coupled to the Ti_{int} donated electrons (Fig. 5.12). The theoretical description of the reactivity of O2 with O_{vac} and OH_b can now be re-examined in the

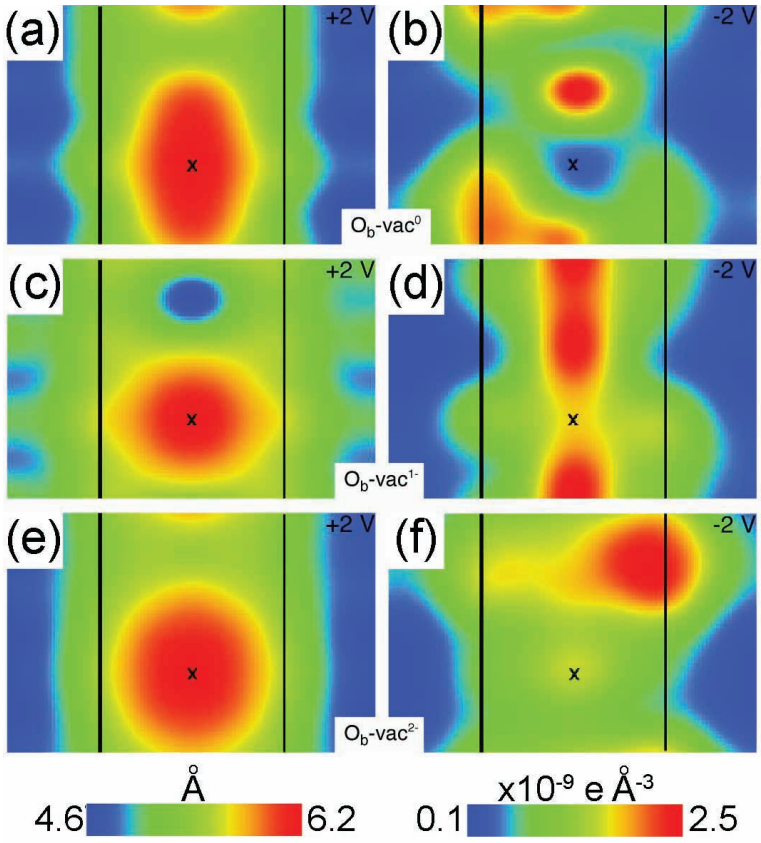

Figure 5.10: Modeled STM and CITS appearance for O_{vac}. Simulated STM images (+2 V, $10^{-7}e\text{Å}^{-3}$) and current maps (-2 V, same height above the surface as from the corresponding left-side topography) are shown for $O_{vac}(0)$ in A and B, for $O_{vac}(1\text{-})$ in C and D, and for $O_{vac}(2\text{-})$ in E and F. Ti5c rows are indicated by black lines and an X marks the O_{vac}. STM simulations were performed with a tip-surface distance of 5 Å[the detailed procedure can be found in ref. [43]].

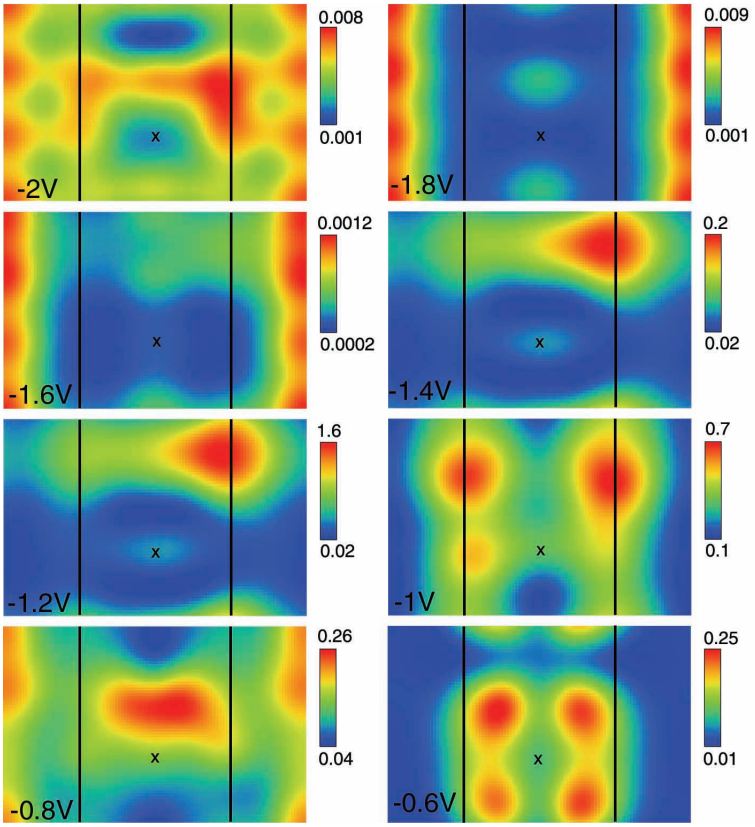

Figure 5.11: O_{vac}(2-) energy selective local density of states maps ($10^{-9}e\text{Å}^{-3}$). Bias windows of ±0.1 V have been considered around the reported bias. Each point of the two-dimensional maps has been evaluated at the same height above the surface as in Fig. 5.3e. Ti_{5c} rows are indicated by the black lines and the O_{vac} is marked withX.

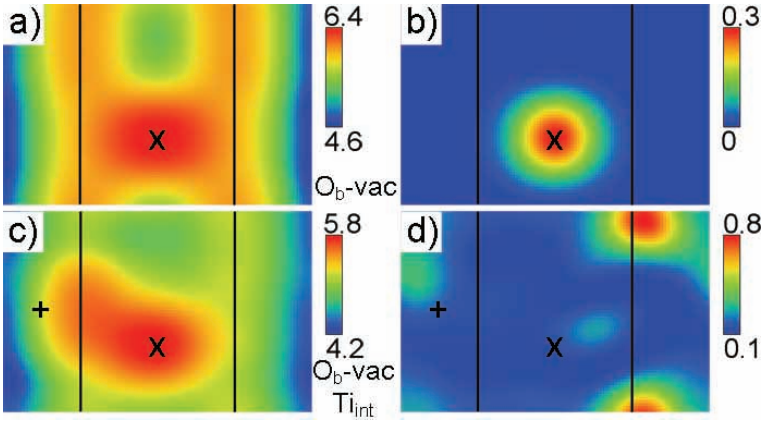

Figure 5.12: Modeled STM and CITS appearance for O_{vac} in the presence of one Ti_{int} between the 4th and 5th tri-layer. Simulated STM images (left: +2V, $10^{-7}e\text{Å}^{-3}$) and LDOS maps (right: -2 V, same height above the surface as from the corresponding left-side topography) are shown for O_{vac} in (a) and (b), for O_{vac}/Ti_{int} in (c) and (d). Heights are reported in Å, LDOS in $10^{-9}e\text{Å}^{-3}$. Ti_{5c} rows are indicated by black lines. (X) marks the O_{vac} site and (+) marks the lateral position of Ti_{int} on the scan areas. The simulated asymmetry for the modeled STM topography of mixed O_{vac}+ Ti_{int} system (c) is less than 15% of the baseline signal. This suggests that Ti_{int} atoms deeper than the fourth tri-layer should hardly be resolved by STM, as confirmed by simulation of one Ti_{int} between the 4th and 5th tri-layers (Fig. 5.15).

presence of Ti_{int}-donated electron surplus. The use of the HSE06 [15] approach allows us the accurate description of the relative energies of molecular H_2O(g), O_2(g), H_2(g). Consequently, the level of accuracy of the calculated chemical potential and thence the grand-canonical formation energies are expected to be sufficiently high to quantitatively address the thermodynamic stability of the considered systems. To match with the STM experiments, the results of the calculations from a simulated temperature of 300 K are used but it is noted that the same trends are found at a large range of temperatures around 300 K. The formation energies of the pertinent adsorbates (H_2O_2, H_2O, O_2H, OH_t, and O_{ad}) are shown in Fig. 5.13 alongside those for O_{vac} and OH_b.

5.8 Energetics

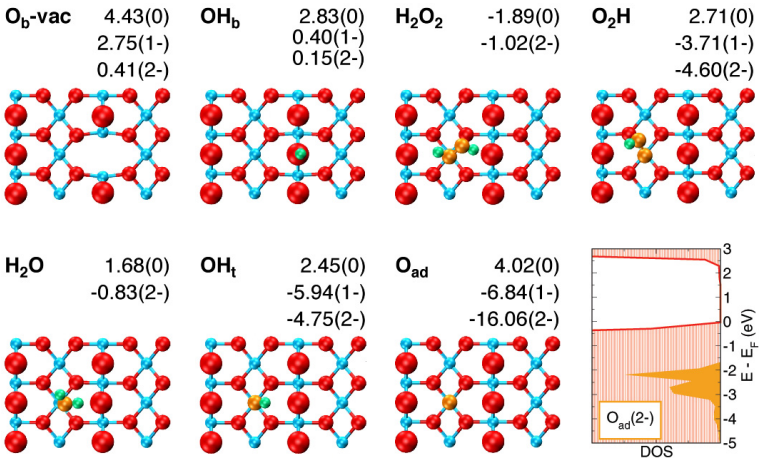

Figure 5.13: Plan view of surface species together with their grand canonical formation free energies (T = 300 K, $P_{O2} = 1.3 \times 10^{-8}$ mbar, $P_{H2O} = 1 \times 10^{-11}$ mbar). Ti is shown blue, lattice O red, O from adsorbates is orange, and H is shown as green. Bridging oxygen (O_b) atoms are shown larger to highlight them. The energies are in eV and the brackets after the energies indicate the electronic charge of the surface species. Geometries are optimized (3×2 super-cell) for the neutral state. The total density of states for the optimized layers (filled red) of O_{ad}(2-)/TiO_2(110) is displayed together with the O_{ad}(2-)-projected density of states (PDOS, filled orange) in the bottom-right panel.

It should be noted here that owing to the necessity of enforcing a compensating background in the simulation cells, only relative energy differences for one specific charge state can be meaningfully compared. To illustrate this point, the energy of H_2O_2(2-) can be compared with O_{ad}(2-) or O_2H(2-) but not with H_2O_2(0) or O_{ad}(0). The calculations indicate that when O2 reacts with O_{vac}(2-) ($\Delta G_f = 0.41$ eV), O_2 dissociates, filling the vacancy and forming one O_{ad}(2-) ($\Delta G_f = -16.06$ eV) with $\Delta G = -16.47$ eV [O_{vac}(2-) + O_2(g) $\rightarrow$ O_{ad}(2-) + O_b], consistent with experiments [5, 7]. As for the reaction of

O_2 with OH_b, it is found that while formation of O_2H(2-) from OH_b(2-) is spontaneous (OH_b(2-) + O_2(g) $\rightarrow$ O_2H(2-) + O_b, $\Delta G = -4.75$ eV), the formation of H_2O_2(2-) ($\Delta G_f = -1.02$ eV) is strongly unfavoured with respect to O_{ad}(2-) ($\Delta G_f = -16.06$ eV). Thus the reaction between O_2 and OH_b must take place via formation of O_2H(2-) which, as with other O_xH_y(2-) intermediates, eventually reacts to form O_{ad}(2-). O_2 acts to scavenge all the initially available electron excess. Consistent with available spectroscopic data (3, 10, 14) where exposure to O2 quenches the BGS, sample oxidation in the calculation is accompanied by removal of Ti^{3+} BGS and localisation of O_{ad}(2-) states in the valence band, VB (Fig. 5.13). Similar results are obtained by simulating one Ti_{int} in

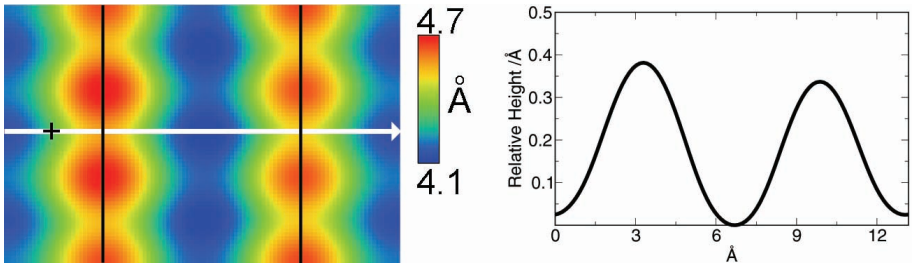

Figure 5.14: Simulated LSDA+U STM image (left: +2V, $10^{-7} e\text{Å}^{-3}$) for a TiO_2(110) 3×2 8 tri-layer slab with one interstitial Ti atom between the 4th and 5th tri-layer (see Fig. 5.6c). Ti_{5c} rows are indicated by black lines. Right panel: modeled STM scanline along the white arrow displayed on the left-side topography. (+) marks the lateral position of Ti_{int} on the scan area.

the presence of two O_{ad}, as shown in Fig. 5.15. No BGS where found associated with the O_{ad}, the charge density contributes to the lattice VB instead. A Bader charge analysis reveals a difference of 10% between the modeled atomic charges of these O_{ad} products (0.9 e) and the bridging O_b atoms (1 e), thus highlighting the same oxide hybridization [40, 41] for both species in the simulation cell.

5.9 Conclusions

After considering the theoretical and experimental evidence of charge accumulation in oxygen vacancies on the Rutile TiO_2 surface due to electron donation from bulk Ti_{int} atoms, a new reactivity mechanism can now be proposed, consistent with the experimental data presented by Henderson et al. [7]. The interaction of dioxygen with O_{vac} leads to dissociation of O_2 on the doubly charged oxygen vacancy leaving one oxygen adatom on top of a five-fold coordinated Ti atom on the surface, in agreement with

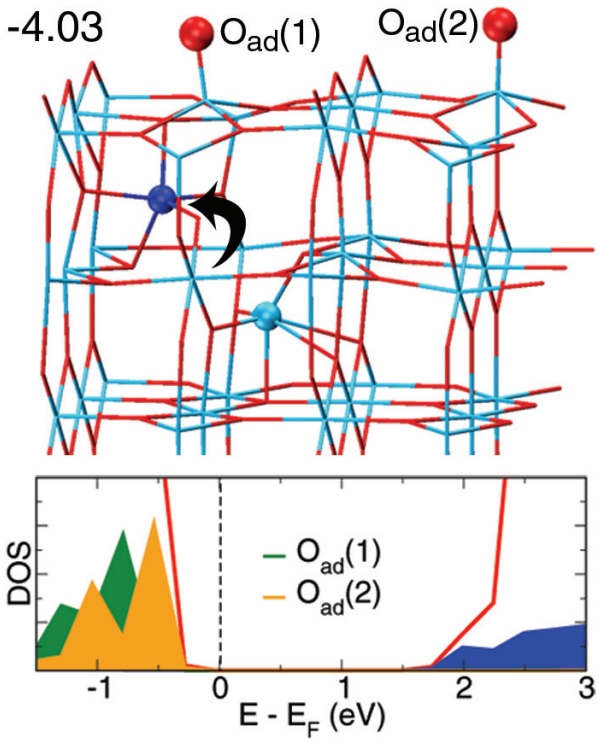

Figure 5.15: Top panel: final geometry and calculated formation energy (eV) for one Ti_{int} between the 2nd and 3rd tri-layers optimised in the presence of a stoichiometric amount of adsorbed atomic oxygen. Blue intersections indicate Ti atoms and red intersections O atoms. O_{ad} atoms are displayed as red spheres. Ti_{int} and the displaced Ti atom are shown as blue and dark blue spheres, respectively. The black arrow highlights the displacement of the newly formed Ti_{int} (dark blue sphere) from its original lattice position. The bottom panel reports the total density of states for the optimized layers (continuous red line) together with the O_{ad}-projected density of states (filled orange and green) and cumulative Ti_{int}-projected (filled blue) density of states.

experiments and other theoretical studies [5, 8, 11]

$$O^{-2}_{vac} + O_2(g) \rightarrow O^{-2}_{ad}, \qquad \Delta G = -16.47eV. \tag{5.9.1}$$

The formation energies for O^{-2}_{vac} and O^{-2}_{ad} are 0.41eV and -16.06eV respectively (Fig. 5.13)[1]. The interaction of O_2 with bridging hydroxyls produced after water dissociation on surface oxygen vacancies proceeds by formation of O_2H^{-2}

$$OH^{-2}_{br} + O_2(g) \rightarrow O_2H^{-2}, \qquad \Delta G = -4.75eV \tag{5.9.2}$$

with O_2H^{-2} eventually decaying to O^{-2}_{ad} [19].

Concluding, in this section, an analysis based on DFT/STM simulations and STM experiments was presented regarding the reactivity of the Rutile TiO_2(110) surface. The gap between the experimental observations originally made by Henderson [7] and theory was discussed and a proposal was made that can bridge this gap. Evidence for charge accumulation in oxygen vacancies, both experimental and theoretical, was presented along with the role of Titanium interstitial atoms as charge donors. By engaging alternative channels of reactivity due to the presence of the extra charge, the interaction of O_2 with OH_{br} is modeled to leave oxygen adatoms O_{ad} on the surface in agreement with the experiment. The results that where discussed in this chapter have been published in the Proceedings of the National Academy of Sciences of the United States of America [44].

5.10 Bibliography

[1] Akira Fujishima and Kenichi Honda. Electrochemical Photolysis of Water at a Semiconductor Electrode. *Nature*, 238(5358):37–38, 1972.

[2] Ulrike Diebold. The Surface Science of Titanium Dioxide. *Surface Science Reports*, 48(5-8):53–229, 2003.

[3] R. Schaub et al.˙Oxygen Vacancies as Active Sites for Water Dissociation on Rutile TiO2(110). *Physical Review Letters*, 87(26):266104, 2001.

[4] Z. Zhang, O. Bondarchuk, Bruce D. Kay, J. M. White, , and Z. Dohnalek. Imaging

[1] It should be noted here that due to the computational methods employed in this study, a comparison between the formation energies of different defects can only be made for the same charged states.

Water Dissociation on TiO2(110): Evidence for Inequivalent Geminate OH Groups. *Journal of Physical Chemistry B*, 110(43):2184021845, 2006.

[5] Oier Bikondoa et al.˙Direct Visualization of Defect-Mediated Dissociation of Water on TiO2(110). *Nature Materials*, 5:189–192, 2006.

[6] C L Pang et al.˙Tailored TiO2(110) Surfaces and Their Reactivity. *Nanotechnology*, 17(21):5397, 2006.

[7] M A Henderson et al.˙Insights Into Photoexited Electron Scavenging Processes on TiO2 Obtained From Studies of the Reaction of O2 With OH Groups Adsorbed at Electronic Defects on TiO2(110). *Journal of Physical Chemistry B*, 107(2):534–545, 2003.

[8] Antonio Tilocca, Cristiana Di Valentin, and Annabella Selloni. O2 Interaction and Reactivity on a Model Hydroxylated Rutile(110) Surface. *Journal of Physical Chemistry B*, 109(44):2096320967, 2005.

[9] Zhenrong Zhang, Yingge Du, Nikolay G. Petrik, Greg A. Kimmel, Igor Lyubinetsky, and Zdenek Dohnlek. Water as a Catalyst: Imaging Reactions of O2 with Partially and Fully Hydroxylated TiO2(110) Surfaces. *Journal of Physical Chemistry C*, 113(5):19081916, 2009.

[10] Y Du et al. Imaging Consecutive Steps of O2 Reaction with Hydroxylated TiO2(110): Identification of HO2 and Terminal OH Intermediates. *Journal of Physical Chemistry C*, 113(2):666671, 2009.

[11] S Wendt et al.˙Oxygen Vacancies on TiO2(110) and Their Interaction With H2O and O2: A Combined High-Resolution STM and DFT Study. *Surface Science*, 598(1-3):226–245, 2005.

[12] S Wendt et al.˙The Role of Interstitial Sites in the Ti3d Defect State in the Band Gap of Titania. *Science*, 320(5884):1755, 2008.

[13] W S Epling and C H F Peden and M A Henderson and U Diebold. Evidence for Oxygen Adatoms on TiO2(110) Resulting from O2 Dissociation at Vacancy Sites. *Surface Science*, 412-413:333–343, 1998.

[14] M D Rasmussen and L M Molina and B Hammer. Adsorption, Diffusion, and

Dissociation of Molecular Oxygen at Detected TiO2(110): A Density Functional Theory Study. *Journal of Chemical Physics*, 120(2):988–997, 2004.

[15] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid Functionals Based on a Screened Coulomb Potential. *Journal of Chemical Physics*, 118(18):8207, 2003.

[16] S D Barrett. Image SXM, 2008.

[17] Cristiana Di Valentin, Gianfranco Pacchioni, and Annabella Selloni. Electronic Structure of Defect States in Hydroxylated and Reduced Rutile TiO2(110) Surfaces. *Physical Review Letters*, 97(16):166803, 2006.

[18] N A Deskins and M Dupuis. Electron Transport by Polaron Hopping in Bulk TiO2: A Density Functional Theory Characterization. *Physical Review B*, 75(19):195212, 2007.

[19] P Kruger et al.˙ Defect States at the TiO2(110) Surface Probed by Resonant Photoelectron Diffraction. *Physical Review Letters*, 100(5):055501, 2008.

[20] M Batzill et al.˙ Variations of the Local Electronic Surface Properties of TiO2(110) Induced by Intrinsic and Extrinsic Defects. *Physical Review B*, 66(23):235401, 2002.

[21] T Minato et al. The Electronic Structure of Oxygen Atom Vacancy and Hydroxyl Impurity Defects on Titanium Dioxide (110) Surface. *Journal of Chemical Physics*, 130:124502, 2009.

[22] G. Kresse and J. Hafner. Ab Initio Molecular Dynamics for Liquid Metals. *Physical Review B*, 47(1):558–561, 1993.

[23] G. Kresse and J. Furthmuller. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Physical Review B*, 54(16):11169–11186, 1996.

[24] P. E. Blochl. Projector Augmented Wave Method. *Physical Review B*, 50(24):17953–17979, 1994.

[25] G. Kresse and D. Joubert. From Ultrasoft Pseudopotentials to the Projector Augmented Wave Method. *Physical Review B*, 59(3):1758–1775, 1999.

[26] John P. Perdew and Yue Wang. Accurate and Simple Analytic Representation

of the Electron-Gas Correlation Energy. *Physical Review B*, 45(23):13244–13249, 1992.

[27] W. A. Hofer and J. Redinger. Scanning Tunneling Microscopy of Binary Alloys: First Principles Calculation of the Current for PtX (100) surfaces. *Surface Science*, 44(1-3):51–61, 2000.

[28] W. A. Hofer. Challenges and Errors: Interpreting High Resolution Images in Scanning Tunneling Microscopy. *Progress in Surface Science*, 71(5-8):147–183, 2003.

[29] J. Tersoff and D. R. Hamann. Theory of the Scanning Tunneling Microscope. *Physical Review B*, 31(2):805–813, 1985.

[30] J. Tersoff and D. R. Hamann. Theory and Application of the Scanning Tunneling Microscope. *Physical Review Letters*, 50(25):1998–2001, 1983.

[31] J. Bardeen. Tunneling from a Many-Particle Point of View. *Physical Review Letters*, 6(2):57–59, 1961.

[32] CJ Calzado, NC Hernandez, and JF Sanz. Effect of On-Site Coulomb Repulsion term U on the Band-Gap States of the Reduced Rutile (110)TiO2 Surface. *Physical Review B*, 77(4):045118, 2008.

[33] Chris G. Van de Welle and Jorg Neugebauer. First-Principles Calculations for Defects and Impurities: Applications to III-Nitrides. *Journal of Applied Physics*, 95(8):3851, 2004.

[34] Chris G. Van de Welle and Jorg Neugebauer. First-Principles Surface Phase Diagram for Hydrogen on GaN Surfaces. *Physical Review Letters*, 88(6):066103, 2002.

[35] John E. Northrup, R. di Felice, and Jorg Neugebauer. Energetics of H and NH2 on GaN(1010) and Implications for the Origin of Nanopipe Defects. *Physical Review B*, 56(8):4325–4328, 1997.

[36] John E. Northrup. Surface Phonon Frequencies and Eigenvectors on Si(111) sqrt[3] x sqrt[3]: Al. *Physical Review B*, 39(2):1434–1437, 1989.

[37] Stephen Gasiorowicz. *Quantum Physics*. Wiley, 2003.

[38] NA Deskins, R Rousseau, and M Dupuis. Localised Electronic States from Surface Hydroxyls and Polarons in TiO2(110). *Journal of Physical Chemistry C*, 113(33):1458314586, 2009.

[39] M A Henderson. A Surface Perspective on Self-Diffusion in Rutile TiO2. *Surface Science*, 419(2-3):174–187, 1999.

[40] R Resta. Physical Chemistry: Charge States in Transition. *Nature*, 453:735, 2008.

[41] H Raebiger, S Lany, and A Zunger. Charge Self-Regulation Upon Chaning the Oxidation State of Transition Metals in Insulators. *Nature*, 453:763–766, 2008.

[42] H Iddir, S Ogut, P Zapol, and N D Browning. Diffusion Mechanisms of Native Point Defects in Rutile TiO2: Ab Initio Total Energy Calculations. *Physical Review B*, 75(7):073203, 2007.

[43] G Teobaldi et al. Modelling STM Images of TiO2(110) from First-Principles: Defects, Water Adsorption and Dissociation Products. *Chemical Physics Letters*, 437(1-3):73–78, 2007.

[44] A C Papageorgiou et al. Electron Traps and their Effects on the Surface Chemistry of TiO2(110). *Proceedings of the National Academy of Sciences of the United States of America*, 107(6):2391, 2010.

Chapter 6

The $TiO_2(011)$ Surface

6.1 Introduction

Due to the high potential of TiO_2 in industrial applications, a substantial amount of effort has been devoted in understanding the exact geometric structure of its various surfaces. Because of its photocatalytic behaviour, the thermodynamically most stable (110) facet has traditionally been the model metal-oxide semiconductor surface and has therefore attracted the majority of the interest of the community. As a result, a significant amount of information for rutile $TiO_2(110)$ has been available, and the surface is believed to be well understood geometrically [1]. Recent research, however, has indicated that the (011) surface can also be active in photo-catalysis, which has lead to increased research activity regarding its properties [2].

In this chapter, the recent developments regarding the structure of the $TiO_2(011)$ surface will be discussed. Then, an analysis based on STM simulations will be provided, that clarifies an ambiguity between simulated and experimentally obtained STM images, presented in the work done by Gong et al. [3]. This analysis effectively extends the work done by Torrelles et al. [4] and Gong et al. [3] in elucidating the exact structure of the geometry of $TiO_2(011)$. In the following section a review will be given on the current status in $TiO_2(011)$ surface structure characterisation process.

6.2 The $TiO_2(011)$ Surface

The bulk terminated $TiO_2(011)(1 \times 1)$ surface, seen in Fig. 6.1(d), consists of exposed two-fold coordinated O atoms, five-fold coordinated Ti atoms with dangling bonds and bulk-like three-fold coordinated O atoms in the valleys [3]. It has never been observed experimentally as the surface atoms form a (2×1) reconstruction instead [5]. Work done by Beck et al. [5] combined a theoretical and experimental study that provided evidence that the $TiO_2(011)$ surface is a stoichiometric (2×1) reconstruction consisting of one-fold coordinated oxygen atoms as seen in Fig. 6.1(a). Atomically-resolved Scanning Tunneling Microscopy images showed a zig-zag pattern, suggesting, along with the results obtained from DFT calculations, that the surface is terminated by singly coordinated oxygen atoms bonded to the underlying titanium atoms through double bonds. The zig-zag pattern as obtained by Gong et al. can be seen in Fig. 6.4(a). The enhanced photocatalytic activity of this surface was attributed in [5] to the presence of those, $Ti = O$ titanyl groups exclusive to $TiO_2(011) - (2 \times 1)$. This structure is therefore usually referred to as the titanyl model. The zig-zag pattern of the bright spots in the high-resolution STM images obtained in [5], was assigned to the doubly bonded oxygen atoms of the titanyl group and was attributed to the dominance of geometric effects since the O atoms protrude from the surface. It has since been pointed out, however, that this proposed surface exhibits a similar energy as the bulk-terminated (1×1) surface [7], questioning therefore its validity [3]. In addition, a more recent study, performed by Kubo et al. [6], using non-contact atomic force microscopy (NC-AFM), STM, and DFT modelling, suggested the rather different structure seen in Fig. 6.1(b). Kubo et al. discarded the surface suggested by Beck et al. as inconsistent with their experimental and theoretical data. It was suggested in this study that the dangling bonds of the Ti atom in the titanyl model proposed in [5], would bond with neighbouring bridging oxygen atoms, inducing large geometrical distortions. Moreover, it was argued that this type of bonding would result in making the surface energetically less stable. The $TiO_2(011) - (2 \times 1)$ structure was proposed instead as it can seen in Fig. 6.1(b), with missing rows of Ti_2O_2 units. It was argued that this surface has a reduced amount of Ti dangling bonds causing therefore its further stabilisation.

The models (a) and (b) presented in Fig. 6.1 [5, 6] have been recently challenged by a different proposal, which can be seen in Fig. 6.1(c). This structure was suggested

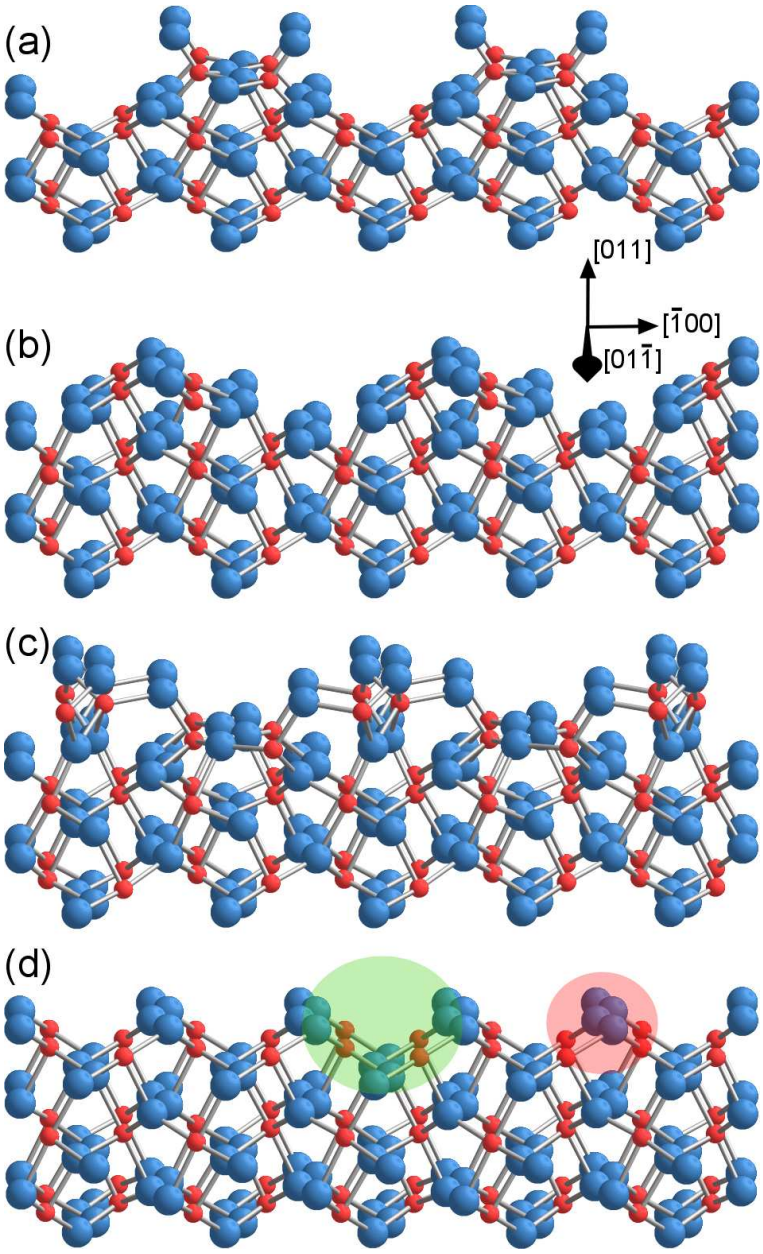

Figure 6.1: Ball and stick models of $TiO_2(011)$. Blue spheres indicate oxygen atoms while red spheres titanium atoms. (a) and (b) indicate structures proposed by Beck et al. [5] and Kubo et al. [6] respectively for the 2×1 reconstruction. (c) indicates the structure proposed Torrelles et al. [4] (d) indicates the atoms lost from the (1×1) surface during formation of Beck et al.'s (green) and Kubo et al.'s (purple) (2×1) surface. Image taken from Torrelles et al. [4]

almost simultaneously by Torrelles et al. [4] and Gong et al. [3]. By employing surface x-ray diffraction (SXRD) and density functional theory modelling, both research groups suggested a geometry consisting of no titanyl groups, requiring only local displacement of atoms, as opposed to the major mass transport required by the previous two proposals. In this structure, the topmost layer terminating the surface consists of twofold coordinated oxygen atoms in a zig-zag arrangement [4] and Fig. 6.2 (bottom) that bind asymmetrically to the underlying five-fold coordinated titanium atoms. This geometry was reported to be substantially more stable compared to the geometries suggested by Beck [5] and Kubo [6] by 1.04 and 2.76 eV per (2×1) unit cell respectively [4].

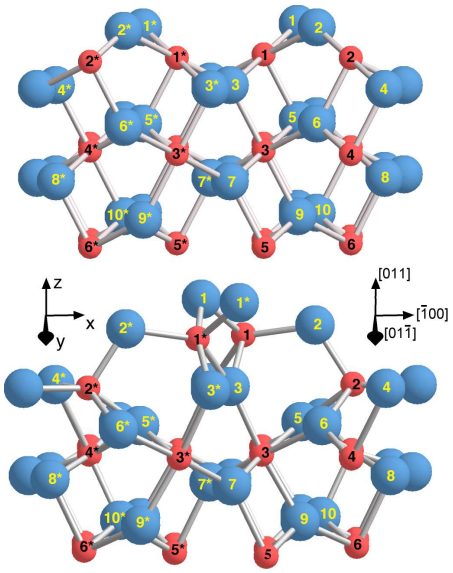

Figure 6.2: Ball and stick models of ideal bulk terminated $TiO_2(011)(1 \times 1)$ (top) and the (2×1) geometry suggested in [4] (bottom). Blue spheres are oxygen atoms, red spheres titanium atoms. Image taken from Torrelles et al.[4]

Fig. 6.3 summarises the geometrical arrangement of the atoms in both the bulk terminated (1×1) and the proposed (2×1) surfaces [4]. The geometrical structure in Fig. 6.3 is the one that has been adopted in the present study. It can be seen that the topmost oxygen atoms O(1) and O(1*) form bonds with different bond lengths with the underlying, five-fold coordinated, Titania, creating the zig-zag rows that can be seen in the STM images reported in [3] and reproduced in Fig. 6.4 (a) and (b).

In the study published by Gong et al. [3], both experimental and simulated STM images were reported for the new proposed surface. The atomically-resolved STM images for the proposed surface Fig. 6.3 and 6.4 (bottom right) were obtained using tunneling parameters of +1.9V, 0.5nA and +1.1V, 0.5nA. They can be seen in Fig. 6.4 (a) and (b) respectively. Simulated images within the Tersoff-Hamann approximation where also

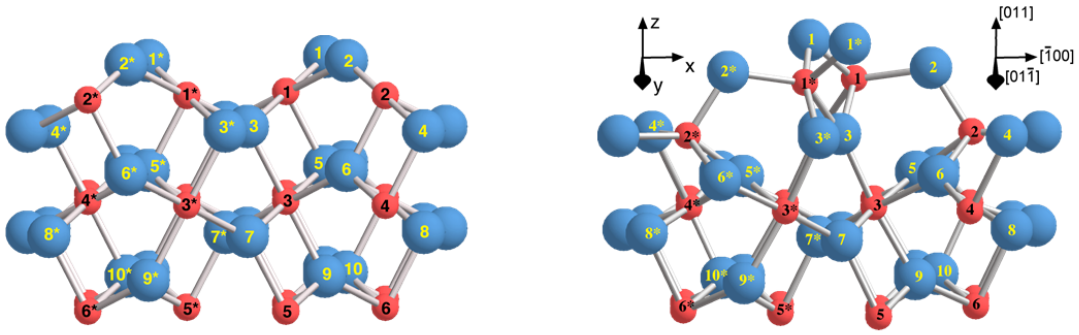

TABLE I. Optimized locations of atoms in the $TiO_2(011)$-(2×1) reconstruction resulting from analysis of the SXRD data and PAW-DFT calculations, expressed as displacements from the bulk-terminated $TiO_2(011)(1 \times 1)$ surface ($\Delta x_{\text{PAW-DFT}}$, $\Delta y_{\text{PAW-DFT}}$, and $\Delta y_{\text{PAW-DFT}}$ have been scaled to take account of the larger theoretical unit cell). (x, y, z) atomic coordinates for the bulk-terminated structure of $TiO_2(011)1 \times 1$ are also listed. Figure 3 provides a key to the identity of the atoms. A positive value for x, y, and z indicates a displacement in the [$\bar{1}$00], [01$\bar{1}$], and [011] directions, respectively.

Atom	(1×1) bulk terminated (x, y, z) coordinates (Å)	Displacement (Å) Δx_{SXRD}: $\Delta x_{\text{PAW-DFT}}$	Δy_{SXRD}: $\Delta y_{\text{PAW-DFT}}$	Δz_{SXRD}: $\Delta z_{\text{PAW-DFT}}$
O(1)	(2.04, 0.00, 0.76)	−2.52 ± 0.02: −2.48	2.90 ± 0.02: 3.00	0.62 ± 0.02: 0.78
O(2)	(2.55, 2.73, 0.76)	−0.15 ± 0.02: −0.20	−0.29 ± 0.02: −0.31	−0.06 ± 0.02: −0.04
Ti(1)	(1.15, 1.56, 0.00)	−0.56 ± 0.01: −0.59	0.32 ± 0.01: 0.32	0.43 ± 0.01: 0.59
Ti(2)	(3.44, 4.28, 0.00)	−0.09 ± 0.02: −0.11	−1.76 ± 0.01: −1.77	−0.80 ± 0.01: −0.74
O(3)	(0.25, 3.11, −0.76)	0.00 ± 0.02: 0.04	0.27 ± 0.02: 0.26	−0.06 ± 0.02: −0.03
O(4)	(4.34, 5.84, −0.76)	−0.15 ± 0.02: −0.15	0.42 ± 0.02: 0.40	−0.01 ± 0.02: −0.06
O(5)	(2.04, 1.61, −1.73)	0.00 ± 0.01: 0.04	−0.04 ± 0.01: −0.04	−0.06 ± 0.02: −0.14
O(6)	(2.55, 4.34, −1.73)	0.03 ± 0.01: 0.03	−0.27 ± 0.02: −0.21	−0.06 ± 0.02: −0.01
Ti(3)	(1.15, 3.16, −2.49)	−0.10 ± 0.01: −0.18	−0.05 ± 0.01: −0.03	−0.06 ± 0.01: −0.03
Ti(4)	(3.44, 5.89, −2.49)	−0.12 ± 0.01: −0.18	−0.27 ± 0.01: −0.27	−0.07 ± 0.01: −0.04
O(7)	(0.25, 4.71, −3.25)	0.04 ± 0.01: 0.03	0.00 ± 0.02: 0.03	−0.05 ± 0.01: −0.01
O(8)	(4.34, 1.98, −3.25)	−0.02 ± 0.01: −0.01	0.07 ± 0.02: 0.10	0.12 ± 0.01: 0.04
O(9)	(2.04, 3.21, −4.22)	−0.02 ± 0.01: −0.03	−0.06 ± 0.02: −0.02	−0.04 ± 0.01: −0.05
O(10)	(2.55, 0.48, −4.22)	0.00 ± 0.01: 0.00	−0.03 ± 0.02: −0.02	0.00 ± 0.01: −0.03
Ti(5)	(1.15, 4.76, −4.98)	0.02 ± 0.01: 0.01	−0.06 ± 0.01: −0.10	−0.11 ± 0.01: −0.12
Ti(6)	(3.44, 2.03, −4.98)	0.01 ± 0.01: 0.00	0.04 ± 0.01: 0.04	0.05 ± 0.01: 0.09

Figure 6.3: Top Left: Bulk terminated $TiO_2(011)(1 \times 1)$ surface. Top Right: The $TiO_2(011)(2 \times 1)$ surface proposed in [4] and [3]. Bottom: Atom coordinates for the two systems as determined by SXRD and DFT, taken from Torrelles et al. [4].

reported, as seen in Fig. 6.4 (bottom left). The STM simulations were performed using a positive bias voltage, with lighter shading in the images corresponding to regions with denser unoccupied states. Images were taken for distances between 2 and 3 Å without any qualitative difference in the simulated features between the different heights [3]. It

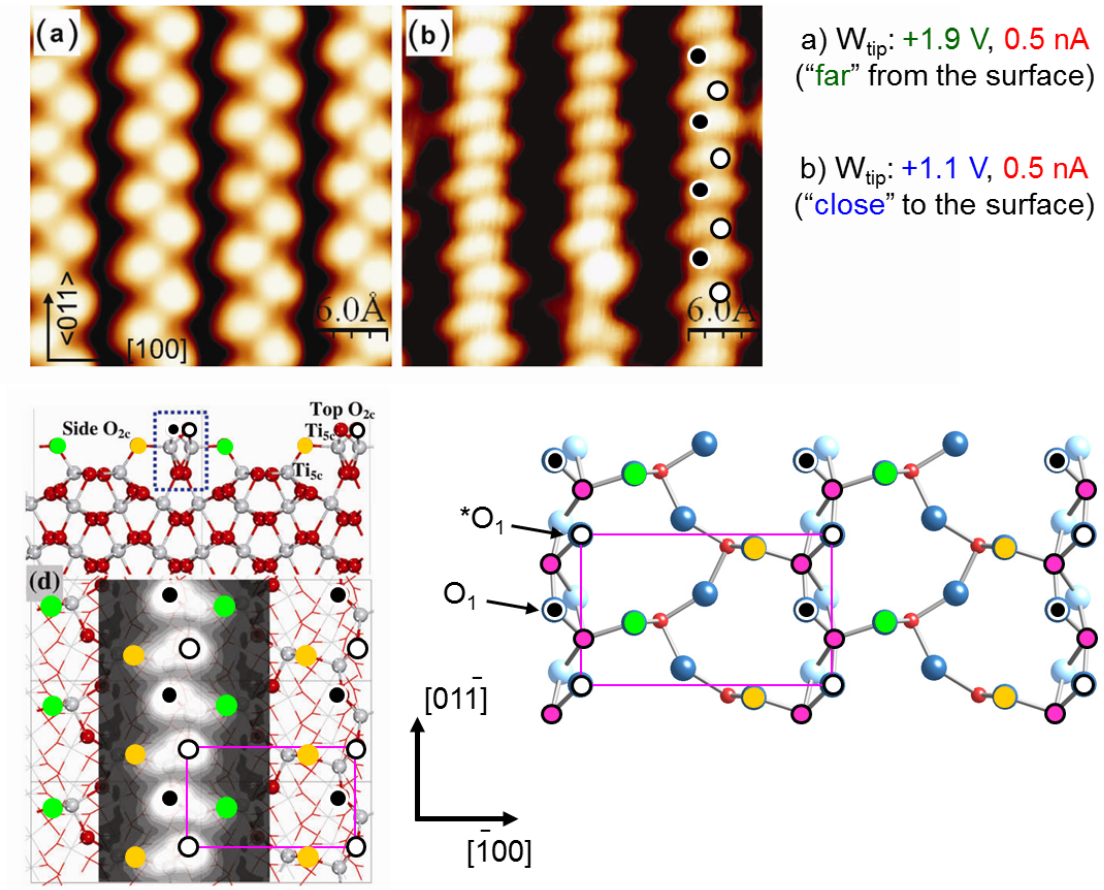

Figure 6.4: Atomically resolved Scanning Tunneling Microscopy images of the $TiO_2(011)(2 \times 1)$ surface [3]. Bias voltage and tunneling current: (a) +1.9V, 0.5nA, (b) +1.1V, 0.5nA. Bottom Left: Side and top view of the $TiO_2(011)(2 \times 1)$ surface. Tersoff-Hamman simulated STM image showing the zig-zag pattern is superimposed to the surface. Bottom right: Top view of the surface. White circles denote the O(1*) atoms while black circles O(1) atoms.

was found that the simulated STM images, as they can be seen in Fig. 6.4 (bottom left), presented a pattern of staggered, bean-like spots, as opposed to the zig-zag pattern of the more round features seen in Fig. 6.4(a). This reflects an inconsistency between the simulated and the experimentally obtained STM images for this surface, with the patterns obtained for the structures by Beck at al. and Kubo et al. more closely resembling the experimental results [3]. It was argued, however, that by adopting different tunneling parameters, as well as different STM tips, a smaller tip-surface separation was possible to be maintained, leading to STM images with zig-zag features as in Fig. 6.4(b)

[3] in a series of subsequent measurements.

It is evident that the STM analysis provided by Gong et al. introduced an ambiguity in the correct interpretation of the experimentally obtained STM images. The relationship between Fig. 6.4 (a) and (b) was not completely clarified, and the single provided simulated STM image 6.4 (bottom left) of the proposed structure [3, 4], along with the similarly simulated images [3] for the structures suggested in [5] and [6] did not uniquely determine the correspondence between the $TiO_2(001)(2 \times 1)$ structure in Fig. 6.1(c) by Torrelles et al. [4] and Gong et al. [3] and the experimentally obtained STM images in Fig. 6.4 (a) and (b). As atomically-resolved STM images were not discussed in [4], the issue has remained open since then. The attempt to reconcile the $TiO_2(001)(2 \times 1)$ structure seen in Fig. 6.1(c) with Fig. 6.4 (a) and (b) will be discussed in the remaining of this chapter.

6.3 Computational Details

Density Functional Theory calculations were performed within a plane-wave framework as implemented in the code VASP [8, 9]. Atoms were described using the projected augmented waves (PAW) method [10, 11]. For geometry optimisations, exchange and correlation were described using the PBE exchange-correlation functional [12] using a (1×2) super-cell slab with eight Ti layers. The upper four layers of the slab were relaxed until all the forces were less than 0.01 $\text{eV}\text{Å}^{-1}$ while the rest of the atoms were kept fixed to their bulk positions. Scanning Tunneling Microscopy simulations were performed within the Tersoff-Hamman approximation using the code bSKAN [13–16]. Wavefunctions were obtained using VASP after performing single-point calculations on the PBE optimised geometries using the HSE06 hybrid exchange-correlation functional [17–20]. The sampling of the Brillouin zone was performed using a $2 \times 3 \times 1$ $(2 \times 2 \times 1)$ Monkhorst-Pack k-point grid [21] with 6 irreducible k-points. The plane-wave energy cutoff was set to 400 eV. LSDA+U simulations where performed using $U_{\text{eff}} = 5.55eV$ as suggested in [22].

6.4 Simulated STM Images of $TiO_2(001)(2 \times 1)$

In this section, the relationship between Fig. 6.4 (a) and (b) as reported by Gong et al. [3] is investigated. The $TiO_2(001)(2 \times 1)$ surface was modeled based on the results reported by Torrelles et al. as seen in Fig. 6.3 [4]. The computational details were as discussed in the previous section. Fig. 6.5 - 6.6 show simulated STM topographies as obtained for the three different DFT frameworks adopted in this study. It can be seen that the simulated STM images are sensitive to the DFT framework adopted due to the differently calculated wavefunctions.

The simulated Tersoff-Hamann STM appearances within the PBE approximation of the $TiO_2(011)(2 \times 1)$ model as seen in Fig. 6.3 and discussed in [4] can be seen in Fig. 6.5. Different tip-surface separations are reported. When studied with PBE-GGA, the calculated contrast close to the surface has the shape of an array of bean-shaped features pinned at O(1)/O(1*) sites, Fig. 6.3 (top left). Simulating the STM images at larger distances from the surface, however, i.e. at lower LDOS iso-contours, the contrast gets progressively blurred. For the greatest tip-surface separation, Fig. 6.3 (middle left), bright stripes are observed running along the O(1*)/O(1) row. These bright stripes are in disagreement with the experimental data recorded by Gong et al. [3] as seen in Fig. 6.4.

In stark contrast with the PBE case, the Tersoff-Hamann simulations within the HSE06 framework indicate a change in the calculated contrast for the surface. This can be seen in Fig. 6.6. While close to the surface (high LDOS iso-contour) an array of bean-shaped features is recovered, pinned at O(1)/O(1*) sites. After increasing the distance of the tip from the surface, the calculated contrast changes. More specifically, the zig-zag pattern of the more rounded features is gradually recovered. As shown in Fig. 6.6 these bright features are now centered in proximity of the under-coordinated (5-fold) Ti(1), Ti(1*) sites. It is important to be noted that the bright features are not centered on Ti(1)/Ti(1*) sites but are actually centered midway the O(1) [O(1*)] and O(2) [O(2*)] atoms. These results suggest that in both cases (bean-shaped and zig-zag features) the STM contrast of TiO2(011)-2x1 is governed by electronic rather than geometric effects. The bright features do not correspond to the geometric position of atoms but rather to the regions where the LDOS is maximal. This can be explained as follows. The TH simulation is a convolution of both O(2p) states and Ti(3d) states. In addition, the O

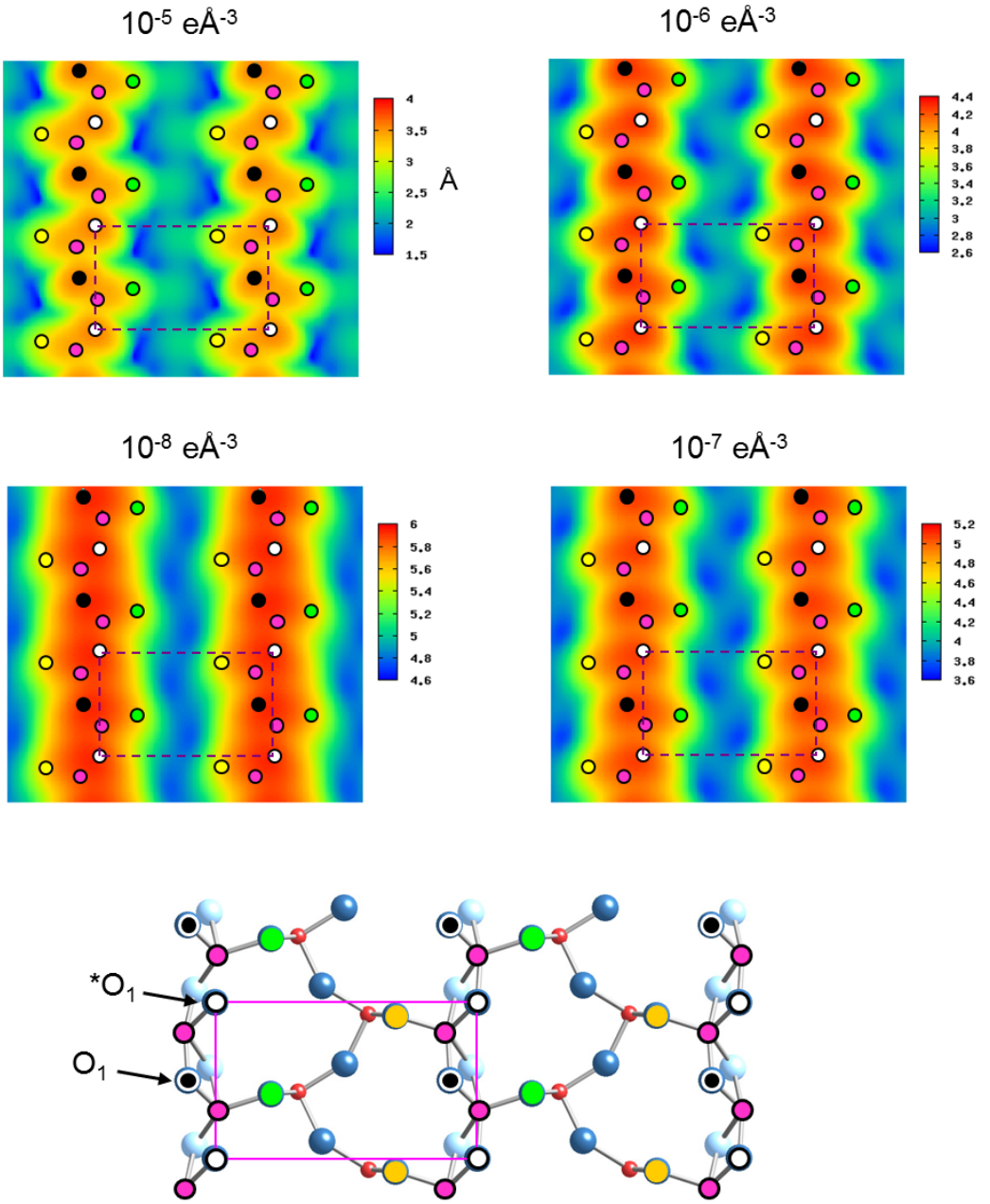

Figure 6.5: Simulated Tersoff-Hamann STM appearance of $TiO_2(011)$ at different heights above the surface. The simulations have been carried out using the calculated GGA-PBE wavefunctions. Simulated bias voltage is +1.1V

atoms are more exposed, geometrically, than Ti atoms, as it can be seen in Fig. 6.2. The decay length of O(2p) electronic states, however, is shorter than that for the Ti(3d) states. The STM contrast for $TiO_2(011)(2 \times 1)$, therefore, depends on an equilibrium between the geometric mismatch in the O (Ti) height, and the different decay lengths of the O(2p) and Ti(3d) states. Thus, provided the description of the O(2p) and Ti(3d) states is accurate enough, it is possible to account for the two experimentally imaged contrasts in Fig. 6.4 on the basis of the surface electronic structure arguments alone.

Since the HSE06 exchange-correlation functional is quite expensive computationally, it was also checked whether the LSDA+U description of the 3d states is enough to recover the experimentally measured change in STM contrast. The same U which was previously shown to suitably describe the TiO2(110) surface, both in terms of chemical reactivity and STM appearance [22, 23], was adopted in this analysis. As it can be seen in Fig. 6.7, it is found that LSDA+U, as HSE06 in Fig. 6.6, yields results in agreement with the experiment (Fig. 6.4). Again, close to the surface the bean-shaped features can be observed, while far from the surface the zig-zag pattern of the more rounded features is recovered. The features are again centered midway the O(1) [O(1*)] and O(2) [O(2*)] atoms. This is an important result as it demonstrates that the required accuracy for the description of Ti(3d) states can be obtained within the LSDA+U framework also. This is crucial for calculations on bigger simulation cells as it provides an encouraging starting point for the simulation of the STM appearance of defects on $TiO_2(011)(2 \times 1)$ at realistic concentrations.

While Fig. 6.5 - 6.7 can provide insights into the real space localisation of the $TiO_2(011)(2 \times 1)$ states in the energy range $[E_f, E_f + 1.1]$ eV, they do not distinguish between the Ti and O states. Such information can, however, be obtained by analysing the (Projected) Density of States for the surface. In all three cases (PBE, HSE06, LSDA+U) it is found that most of the contributions to the Conduction Band edge come from the Ti states. The O contribution however is not zero. Thus, the more exposed position of O(1), O(1*) atoms, counterbalances the larger Ti PDOS and the Ti(3d) decay length, making it possible to measure bright features on O atoms rather than Ti atoms. It should be noted that the GGA Ti(PDOS)/O(PDOS) ratio is different compared to the ratios obtained from the HSE06 and LSDA+U simulations. It is confirmed therefore, that, once again, an accurate description of the structure, defined in Fig. 6.3, energy (Fig. 6.8), and real-space localisation (Fig. 6.6 - 6.7) is crucial for the correct modelling of the STM appearance

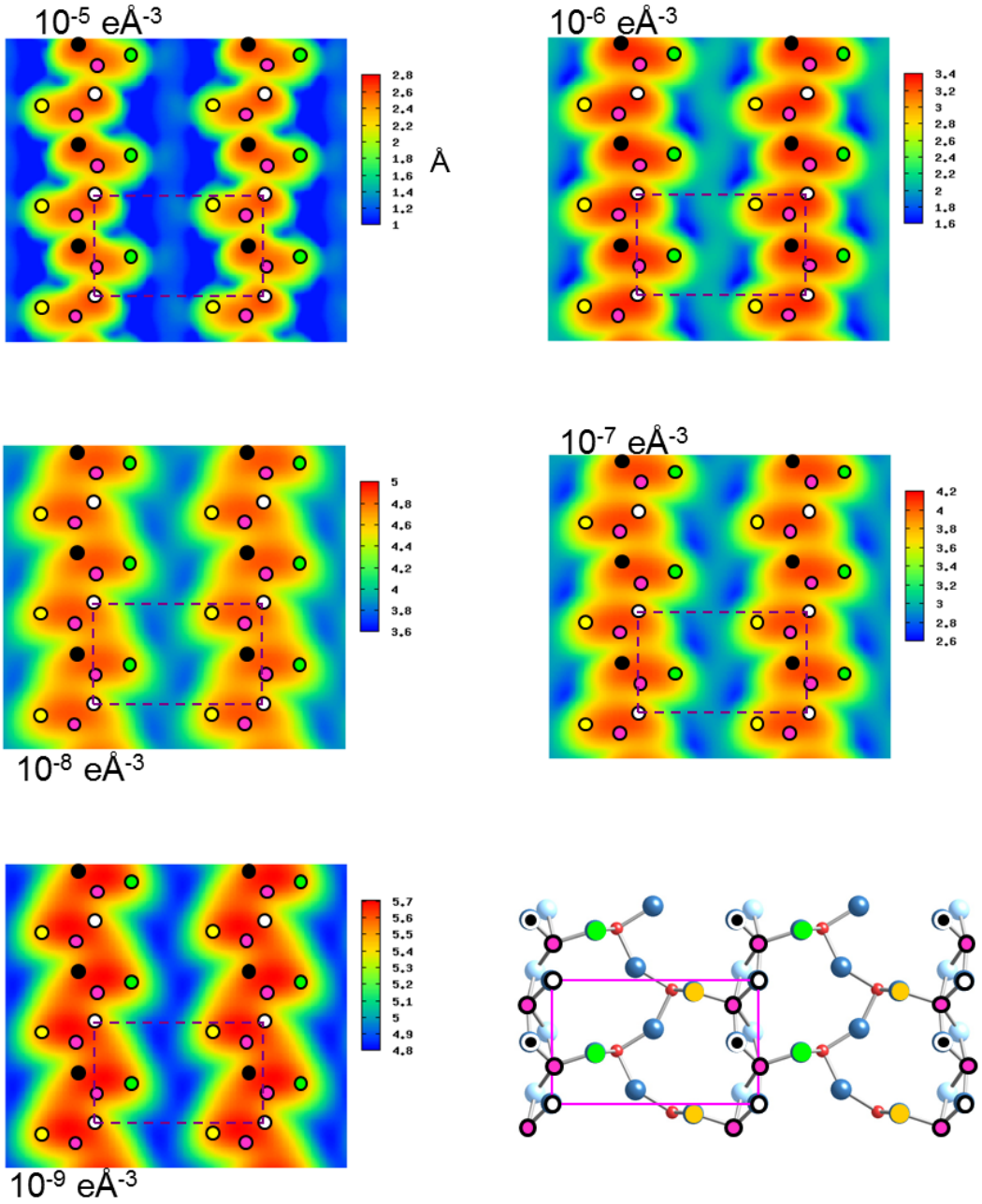

Figure 6.6: Simulated Tersoff-Hamann STM appearance of TiO2(011) at different height above the surface. The simulations have been carried out using the calculated HSE06 wavefunctions. Simulated bias voltage is +1.1V

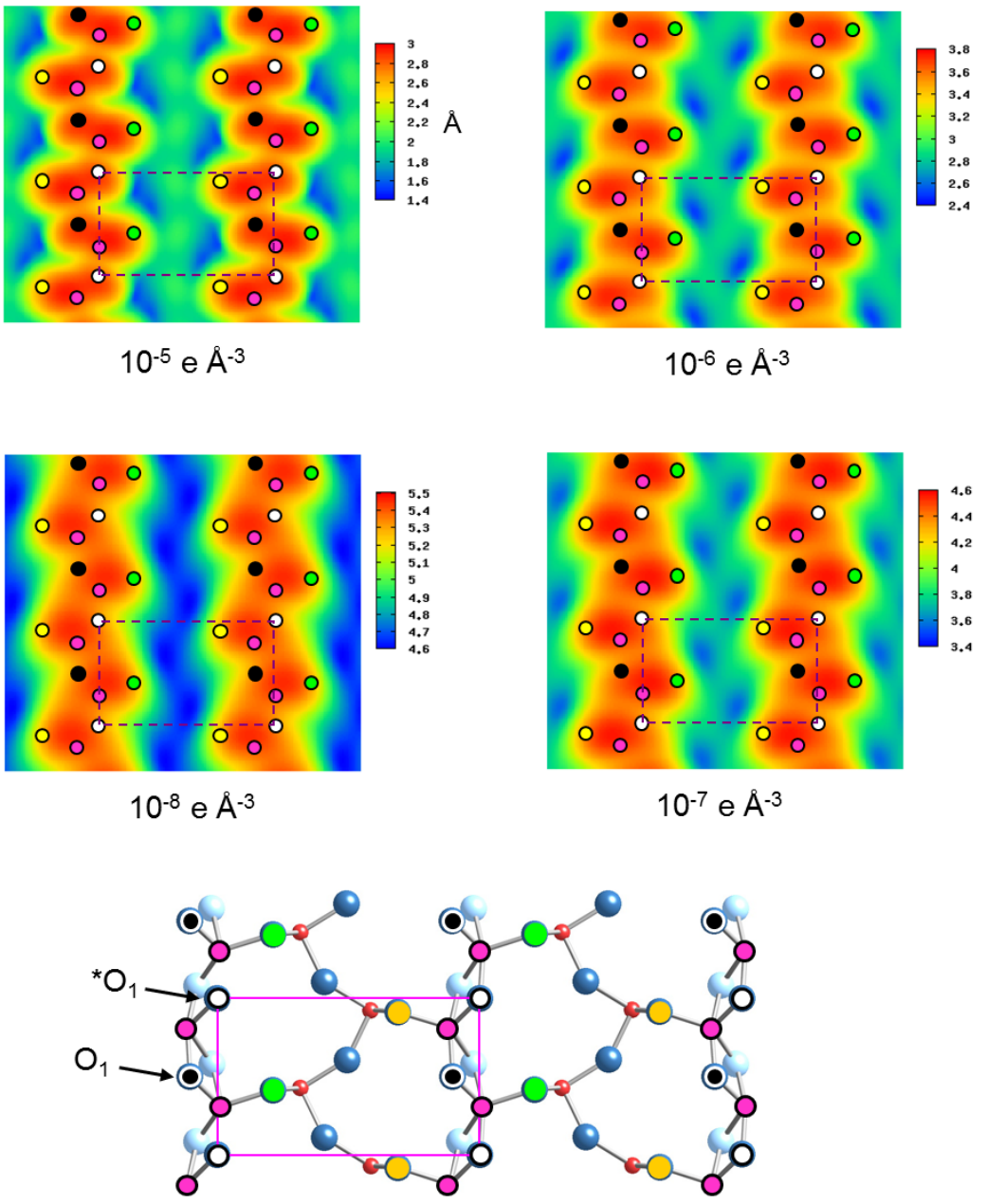

Figure 6.7: Simulated Tersoff-Hamann STM appearance of TiO2(011) at different height above the surface. The simulations have been carried out using the calculated LSDA+U wavefunctions. Simulated bias voltage is +1.1V

of the surface. It could be speculated, at this point, that the same conclusion might apply to STM imaging of other O-terminated Titania surfaces [1] and interfaces [24].

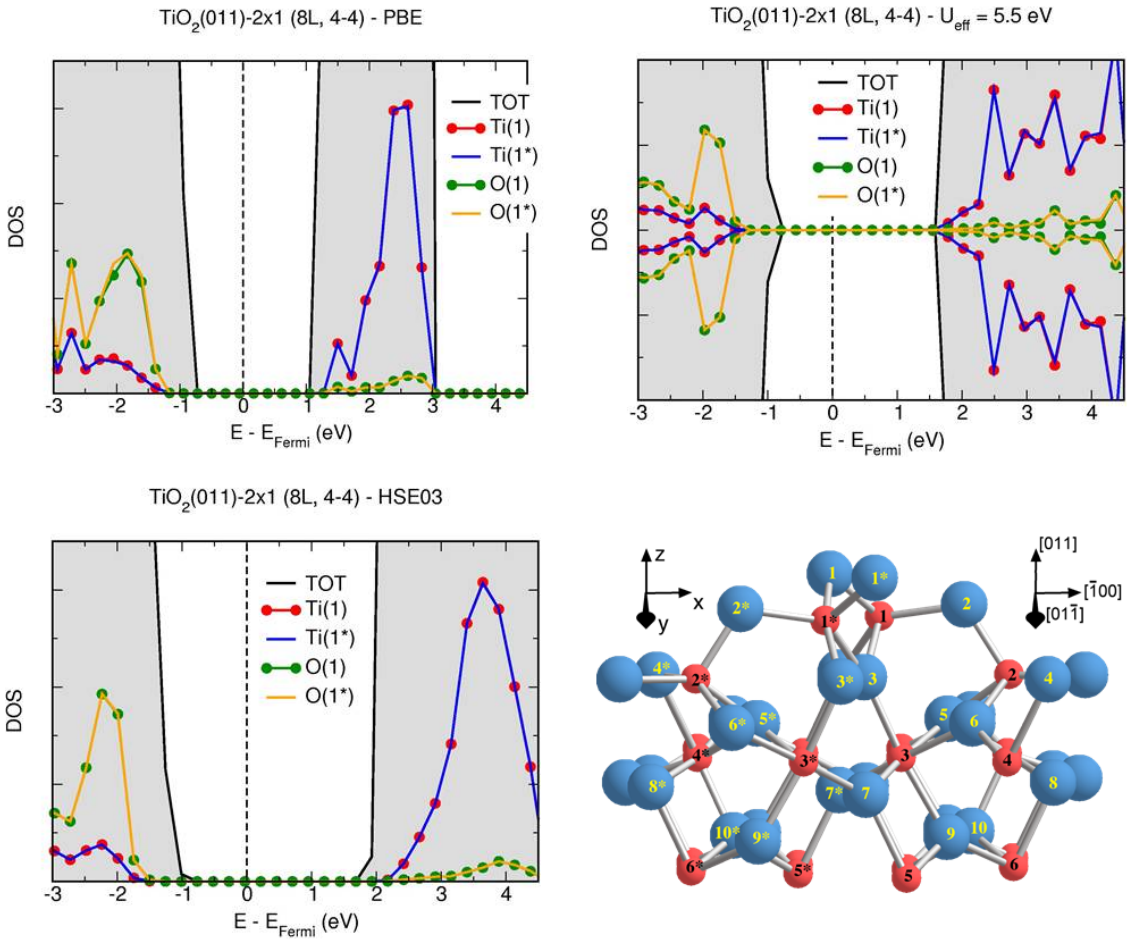

Figure 6.8: Calculated Density of States (DOS, filled gray) and Projected Density of States (PDOS) on Ti(1) (red), Ti(1*) (blue), O(1) (green), O(1*) (yellow) atoms at GGA-PBE (top-left), LSDA+U (top-right), and HSE06 (bottom-left) level. A schematic model of the surface, together with the adopted atomic labeling, is shown in the bottom-right panel.

6.5 Conclusions

The recent advances in understanding the geometrical structure of the $TiO_2(011) - (2 \times 1)$ surface open the possibility of studying its electronic structure along with its reactivity and the role of defects in catalysis and other surface chemical processes [3, 4]. In this study, both pure and hybrid Density Functional Theory was employed to further elucidate the exact structure of the surface. STM simulations of the surface structure suggested by Torrelles et al. [4] and Gong et al. [3] were performed in order to clarify the different STM contrasts observed in [3]. When performing the STM simulations

using wavefunctions obtained from PBE-DFT, an increase in the simulated distance between the tip and the surface lead to a blurred image without any additional information on the underlying structure. Using LSDA+U and the HSE hybrid exchange-correlation functional, however, it was found that the differences in contrast between 6.4 (a) and (b) are due to the change in the surface electronic structure as sampled by the STM tip for different tip-surface separations, see also Fig. 6.6, and not due to a different geometrical arrangement of the surface atoms between the two scanned areas. This is summarised in Fig. 6.9. These results further confirm the validity of the $TiO_2(011)(2 \times 1)$ structure

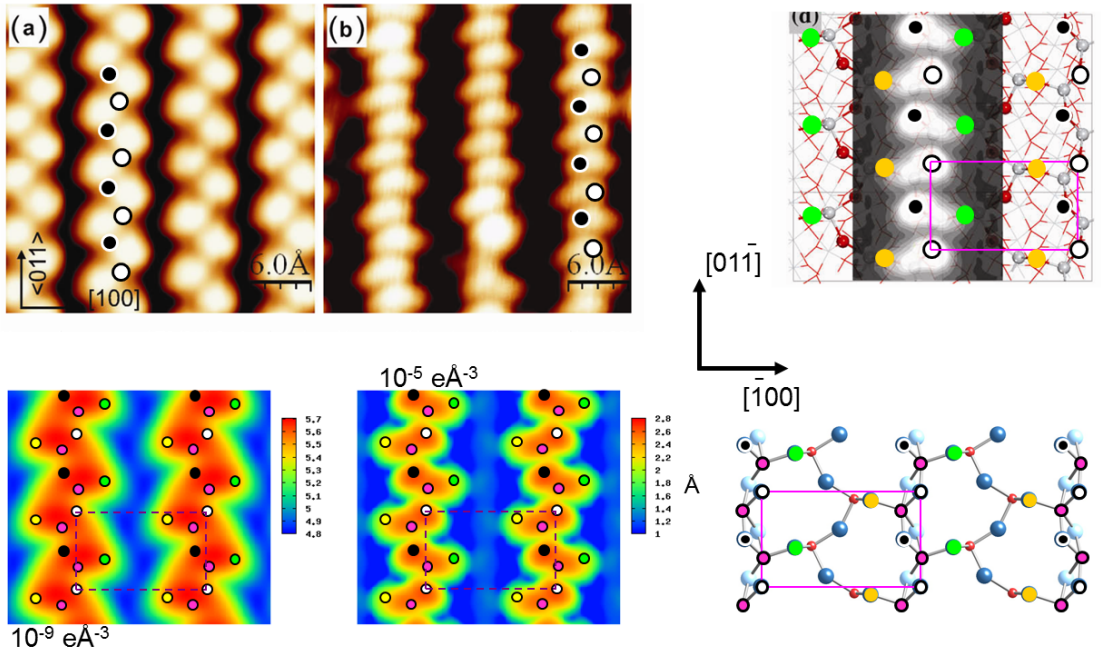

Figure 6.9: $TiO_2(011)$ Results Summary: Top Left: STM Image obtained for the $TiO_2(011)$ surface [3]. (a) Zig-zag pattern far from the surface. (b) Bean-like pattern close to the surface. Top Right: PBE-based simulation within Tersoff-Hamann, close to the surface [3]. Bottom Left: HSE-based simulation within Tersoff-Hamann. Both zig-zag and bean-like patterns are reproduced. Bottom Right: Top view of the $TiO_2(011)$ surface suggested in [3, 4].

proposed by Torrelles et al. and Gong et al. in [4] and [3] respectively.

The understanding of the exact structure of the rutile $TiO_2(011)$ surface and the clarification of the ambiguities appearing in the literature regarding the atomically-resolved STM images of the surface, is the first step towards understanding the electronic structure of the surface. This can lead, as is the case with the rutile $TiO_2(110)$, to a better understanding of the surface's reactivity, the electronic structure of defects, as well as its role in heterogeneous catalysis, with significant fundamental and technological im-

plications.

6.6 Bibliography

[1] Ulrike Diebold. The Surface Science of Titanium Dioxide. *Surface Science Reports*, 48(5-8):53–229, 2003.

[2] Teruhisa Ohno, Koji Sarukawa, and Michio Matsumura. Crystal Faces of Rutile and Anatase TiO2 Particles and Their Roles in Photocatalytic Reactions. *New Journal of Chemistry*, 26(9):1167–1170, 2002.

[3] Xue-Qing Gong et al.˙The 2x1 Reconstruction of the Rutile TiO2(011) Surface: A Combined Density Functional Theory, X-ray Diffraction, and Scanning Tunneling Microscopy Study. *Surface Science*, 603(1):138–144, 2009.

[4] X. Torrelles et al.˙Geometric Structure of TiO2(011)(2x1). *Physical Review Letters*, 101(18):185501, 2008.

[5] T. J. Beck et al.˙Surface Structure of TiO2(011)-(2x1). *Physical Review Letters*, 93(3):036104, 2004.

[6] Toshitaka Kubo, Hideo Orita, and Hisakazu Nozoye. Surface Structures of Rutile TiO2(011). *Journal of the American Chemical Society*, 129(34):1047410478, 2007.

[7] Olga Dulub, Cristiana Di Valentin, Annabella Selloni, and Ulrike Diebold. Structure, Defects, and Impurities at the Rutile TiO2(011)(2x1) Surface: A Scanning Tunneling Microscopy Study. *Surface Science*, 600(19):4407–4417, 2007.

[8] G. Kresse and J. Hafner. Ab Initio Molecular Dynamics for Liquid Metals. *Physical Review B*, 47(1):558–561, 1993.

[9] G. Kresse and J. Furthmuller. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Physical Review B*, 54(16):11169–11186, 1996.

[10] P. E. Blochl. Projector Augmented Wave Method. *Physical Review B*, 50(24):17953–17979, 1994.

[11] G. Kresse and D. Joubert. From Ultrasoft Pseudopotentials to the Projector Augmented Wave Method. *Physical Review B*, 59(3):1758–1775, 1999.

[12] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Physical Review Letters*, 77(18):3865–3868, 1996.

[13] W. A. Hofer and J. Redinger. Scanning Tunneling Microscopy of Binary Alloys: First Principles Calculation of the Current for PtX (100) surfaces. *Surface Science*, 44(1-3):51–61, 2000.

[14] W. A. Hofer. Challenges and Errors: Interpreting High Resolution Images in Scanning Tunneling Microscopy. *Progress in Surface Science*, 71(5-8):147–183, 2003.

[15] J. Tersoff and D. R. Hamann. Theory of the Scanning Tunneling Microscope. *Physical Review B*, 31(2):805–813, 1985.

[16] J. Tersoff and D. R. Hamann. Theory and Application of the Scanning Tunneling Microscope. *Physical Review Letters*, 50(25):1998–2001, 1983.

[17] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid Functionals Based on a Screened Coulomb Potential. *Journal of Chemical Physics*, 118(18):8207, 2003.

[18] Jochen Heyd and Gustavo E. Scuseria. Efficient Hybrid Density Functional Calculations in Solids: Assessment of the Heyd-Scuseria-Ernzerhof Screened Coulomb Hybrid Functional. *Journal of Chemical Physics*, 121(3):1187, 2004.

[19] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Erratum: Hybrid Functionals Based on a Screened Coulomb Potential. *Journal of Chemical Physics*, 124:219906–219906–1, 2006.

[20] Cristiana Di Valentin, Gianfranco Pacchioni, and Annabella Selloni. Electronic Structure of Defect States in Hydroxylated and Reduced Rutile TiO2(110) Surfaces. *Physical Review Letters*, 97(16):166803, 2006.

[21] Hendrik J. Monkhorst and James D. Pack. Special Points for Brillouin Zone Integrations. *Physical Review B*, 13(12):5188–5192, 1976.

[22] CJ Calzado, NC Hernandez, and JF Sanz. Effect of On-Site Coulomb Repulsion term U on the Band-Gap States of the Reduced Rutile (110)TiO2 Surface. *Physical Review B*, 77(4):045118, 2008.

[23] A C Papageorgiou et al. Electron Traps and their Effects on the Surface Chemistry

of TiO2(110). *Proceedings of the National Academy of Sciences of the United States of America*, 107(6):2391, 2010.

[24] Anthoula Chrysa Papageorgiou et al. Growth and Reactivity of Titanium Oxide Ultrathin Films on Ni(110). *The Journal of Physical Chemistry C*, 111(21):7704–7710, 2007.

Chapter 7

Summary and Future Work

7.1 Discussion

Since the discovery of the transistor and the advances in solid state physics, an unprecedented increase in available computing power has been observed. From computer workstations[1] to massively parallel supercomputers, and from general purpose CPUs to specialised processing units, the availability and accessibility of computational resources has been substantially commoditised. The enhanced capability of performing extremely complex numerical calculations is now allowing the study of a wide variety of scientific problems. Computers now assist in deepening our understanding in as diverse fields as weather modeling and global climate change, vaccines design, financial markets, radiation damage, materials modeling, and many more. Although hardware evolution has allowed the realisation of such advances, it has traditionally been driven by the demand from the end users, both from the mainstream consumers' market as well as from the more specialised governmental, academic and corporate sectors. The scientific community, and more particularly physicists, have traditionally been one of the major users, as well as innovators, of computer technology. It is natural, therefore, that a variety of numerical and computational techniques have been devised to facilitate research in analytically hard-to-track problems [1, 2].

Density Functional Theory [3] has been one of the most prominent computational theo-

[1] It is slowly becoming possible to achieve Teraflop power in single workstations. This is happening by utilising double-precision enabled graphics cards and their associated APIs like CUDA and OpenCL. Although this is a field of active research, with major advances in software and hardware design announced almost every year, numerical algorithms have been ported and are slowly being adopted by academic and corporate institutions for production calculations.

ries for electronic structure calculations. It is today used for analysing the properties of a wide range of materials. Despite the initial and persisting shortcomings, the relative computational efficiency compared to other methods and the advances in approximations adopted, like, for example, in the description of exchange and correlation, have made DFT an extremely successful computational scheme. The first chapter of the thesis was dedicated to providing a brief overview of the formalism and the mechanics of DFT. The early steps where first discussed, then the breakthroughs due to Hohenberg, Kohn and Sham, to recent developments in exchange-correlation functionals, pseudopotentials, and implementations were surveyed.

In the second chapter of the thesis, the theoretical formulation of the Scanning Tunneling Microscope was discussed. Since its discovery in 1981, the STM has emerged as the prominent tool for atomic resolution imaging of surfaces. The substantial effort devoted in understanding the electron transport problem in metal-insulator-metal junctions, assisted in obtaining perturbative and exact formulations for the tunneling current flowing between the tip of the STM and the surface under study. It is now therefore possible to combine DFT calculations and STM simulations to obtain simulated STM images for the surface. From the less expensive Tersoff-Hamann framework, to the complete perturbative treatment using non-equilibrium Green's functions, the tunneling current can be simulated and therefore provide further insights in the quantum mechanical interactions between the tip and the surface. By effectively assisting in the interpretation of the STM images obtained from the experimental methods, the computational implementation of the theory behind the STM has allowed for more robust and efficient studies of surfaces, their defects, and their role in surface processes like catalysis among others. Recent experimental and theoretical advances have also allowed for a departure from plain constant-height and constant-current topographies. An expansion to more sophisticated techniques involving inelastic scattering, by the coupling of the electrons to vibrational modes introduced in the barrier region by adsorbed molecules, provides an additional tool for analysing the surface.

In this thesis, three model metal-oxide semiconducting systems were investigated using the techniques described in the previous paragraphs. Continuing from chapters 2 and 3, chapter 4 focused on hydroxyl vacancies on aluminosilicate, $(Al_2SiO_7H_4)_{24}$, and aluminogermanate, $(Al_2GeO_7H_4)_{36}$, nanotubes. The impact of the defects on the geometry, electronic structure, and mechanical properties of the nanotubes was investigated,

in view of their technological applications. The defects, when introduced on both sides of the tube walls, lead to occupied and empty states in the band gap. Those states were found to be highly localised both in energy and in real space. Different magnetisation states were also found, depending on both the chemical composition and the specific side with respect to the tube cavity. The defect-induced perturbations to the pristine electronic structure were related to the electrostatic polarisation across the tube walls and the ensuing change in Lewis acid-base reactivity. A general approach towards a quantitative evaluation of both the polarisation across the tube walls, and the tube excluded volume, was also proposed and discussed on an electrostatic basis [4]. Further study of those systems and their associated hydroxyl vacancies can be pursuit, by studying from first-principles processes like dehydration, dehydroxylation, and rehydroxylation, in line with recently acquired experimental data [5]. On a slightly different route, and representing a cross-over between the work presented in this chapter and the subsequent ones, the properties of defects on TiO_2-based nanotubes [6] could be investigated using both pure, hybrid and L(S)DA+U based DFT.

In chapter 5 of the thesis, an analysis based on DFT/STM simulations and STM experiments was presented, regarding the reactivity of the rutile $TiO_2(110)$ surface. The gap between the experimental observations originally made by Henderson [7] and theory was discussed and a proposal was made that bridges this gap. Evidence for charge accumulation in oxygen vacancies, both experimental and theoretical, was presented, along with the role of Titanium interstitial atoms as charge donors. It was argued that the presence of the extra charge, engages alternative channels of reactivity that allow the interaction of molecular oxygen with surface hydroxyls, formed from water dissociation on oxygen vacancies, to be tracked. This interaction between O_2 and OH_{br} was modeled to favour a condensation reaction forming water and oxygen adatoms O_{ad}, in agreement with the experiment [8]. This new picture that emerged regarding the nature of the oxygen vacancy in $TiO_2(110)$, can have significant implications in other areas of $TiO_2(110)$ research. A brief overview of the research activity in $TiO_2(110)$ supported Au nanoparticles is given below. It will then become apparent how the subject can be re-approached using the results that were derived in chapter 5.

Although it is well known that bulk gold is chemically inert [9], it has been suggested that when gold nanoparticles are dispersed on metal oxides, they become chemically active and can assist in processes like catalysis and oxidation [10–14]. Lopez et. al.

argued that anchoring of gold nanoparticles on titanium dioxide does not take place on a stoichiometric $TiO_2(110)$ surface and that it is necessary for oxygen vacancies, or other defects like step edges, to exist for adsorption to happen [15–17]. This confirmed Wahlstrom et. al. who observed that the number of vacancies on the surface reduces after Au dispersion. In the same work of Wahlstrom et. al. it was also argued that each vacancy can hold an average number of 3 Au atoms per vacancy and that growth of Au clusters happen by vacancy diffusion and Au-O_{vac} coalescing [18]. Pillay and Hwang however suggested that it is possible for gold to nucleate also on 5-fold coordinated Ti sites provided that the Au deposition rate is higher than the surface diffusion rate [19]. The high catalytic activity of Au nanoclusters of size of approximately 3nm has been attributed to charge transfer from the TiO_2 support to the cluster [17, 20]. More recently, Matthey et. al. studied the nucleation of small gold clusters on three different phases of Rutile $TiO_2(110)$ namely the reduced, hydroxylated and oxidized surfaces where it was found that gold bonds to the support more strongly when the surface oxygen vacancies are filled with oxygen atoms through O_2 dissociation on O_{vac} leading to the presence of oxygen adatoms above 5-fold coordinated Ti atoms. They argued that the stabilization of Au nanoparticles is due to Au-O-Ti bonds and that reduced surfaces of metal-oxides offer an O-rich and electron-rich environment that enhances the catalytic properties of the dispersed nanoclusters [21]. Further theoretical studies using pure and hybrid Density Functional Theory supplemented by Scanning Tunneling Microscopy and Spectroscopy experiments and simulations can be performed on TiO_2 supported Au nanoclusters, with the aim of understanding the effect of the recently reported charge accumulation on the O_{vac} [8] on Au nucleation, the role of charge in the observed catalytic properties of Au, and how it affects the outcome of STM and STS experiments.

The final chapter of the thesis focused on another surface of TiO_2. The (011) surface has recently attracted attention due to its enhanced photocatalytic activity. Several studies have been published recommending different geometrical structures for the surface [22–25]. Those structures were discussed, and an STM analysis was performed based on different exchange-correlation functionals within the Density Functional Theory framework. By using the PBE, LSDA+U and the hybrid HSE exchange-correlation functionals, it was found that the previously published STM images [23] are consistent with the structure proposed in [23, 24]. The differences in contrast between the

published images in [23] were found to be due to the change in the surface electronic structure as sampled by the STM tip for different tip-surface separations, and not due to a different geometrical arrangement of the surface atoms between the two scanned areas [26]. Since the "brookite (001)-like" structure of the TiO2(011)(2 × 1) surface has now been elucidated, it is now possible to further study the properties of this surface. The possibilities are literally endless, with the first steps to be the study of oxygen vacancies and their role in TiO_2(011) reactivity using DFT and STM. H_2O and O_2 dissociation could then be tracked, given the usual experimental conditions in the laboratory. In addition, similarly to TiO_2(110) and the work presented in chapter 5, the presence of extra charge from Ti donors could also be considered.

7.2 Bibliography

[1] Tao Pang. *An Introduction to Computational Physics*. Cambridge University Press, 2010.

[2] J. M. Thijssen. *Computational Physics*. Cambridge University Press, 1999.

[3] Robert G. Parr and Weitao Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press, 1989.

[4] Gilberto Teobaldi et al. . Hydroxyl Vacancies in Single-Walled Aluminosilicate and Aluminogermanate Nanotubes. *Journal of Physics: Condensed Matter*, 21(19):195301, 2009.

[5] Dun-Yen Kang et al. Dehydration, Dehydroxylation, and Rehydroxylation of Single-Walled Aluminosilicate Nanotubes. *ACS Nano*, 4(8):4897–4907, 2010.

[6] Qiang qiang Meng et al. Nanotubes from Rutile TiO2 (110) Sheets: Formation and Properties. *The Journal of Physical Chemistry C*, 114(20):9251–9256, 2010.

[7] M A Henderson et al.˙Insights Into Photoexited Electron Scavenging Processes on TiO2 Obtained From Studies of the Reaction of O2 With OH Groups Adsorbed at Electronic Defects on TiO2(110). *Journal of Physical Chemistry B*, 107(2):534–545, 2003.

[8] A C Papageorgiou et al. Electron Traps and their Effects on the Surface Chemistry

of TiO2(110). *Proceedings of the National Academy of Sciences of the United States of America*, 107(6):2391, 2010.

[9] B. Hammer and J. K. Norskov. Why Gold is the Noblest of all the Metals. *Nature*, 376(6537):238–240, 1995.

[10] Masatake Haruta. Size- and Support-Dependency in the Catalysis of Gold. *Catalysis Today*, 36(1):153–166, 1997.

[11] Masatake Haruta. When Gold is not Noble: Catalysis by Nanoparticles. *The Chemical Record*, 3(2):75–87, 2003.

[12] M. Haruta, T. Kobayashi, N. Yamada, and S. Iijima. Gold Catalysts Prepared by Coprecipitation for Low-Temperature Oxidation of Hydrogen and Carbon Monoxide. *Journal of Catalysis*, 115(2):301–309, 1981.

[13] M. Valden, X. Lai, and D. W. Goodman. Onset of Catalytic Activity of Gold Clusters on Titania with the Appearance of Nonmetallic Properties. *Science*, 281:1647–1650, 1998.

[14] M. Valden, S. Pak, X. Lai, and D. W. Goodman. Structure Sensitivity of CO Oxidation Over Model Au/TiO2 Catalysts. *Catalysis Letters*, 56(1):7–10, 1998.

[15] N. Lopez et al. The Adhesion and Shape of Nanosized Au Particles in a Au/TiO2 Catalyst. *Journal of Catalysis*, 225(1):86–94, 2004.

[16] T. Okazawa, M. Kohyama, and Y. Kido. Electronic Properties of Au Nano-Particles Supported on Stoichiometric and Reduced TiO2(110) Substrates. *Surface Science*, 600(19):4430–4437, 2006.

[17] K. Okazaki, S. Ichikawa, Y. Maeda, M. Haruta, and M. Kohyama. Electronic Structures of Au Supported on TiO2. *Applied Catalysis A: General*, 291(1-2):45–54, 2005.

[18] E. Wahlstrom et al. Bonding of Gold Nanoclusters to Oxygen Vacancies on Rutile TiO2(110). *Physical Review Letters*, 90(2):026101, 2003.

[19] Devina Pillay and Gyeong S. Hwang. Growth and Structure of Small Gold Particles on Rutile TiO2(110). *Physical Review B*, 72(20):205422, 2005.

[20] Taketoshi Minato et al. Investigation of the Electronic Interaction Between

TiO2(110) Surfaces and Au Clusters by PES and STM. *Surface Science*, 566-568(2):1012–1017, 2004.

[21] D. Matthey et al. Enhanced Bonding of Gold Nanoparticles on Oxidized TiO2(110). *Science*, 315:1692–1696, 2007.

[22] T. J. Beck et al.˙ Surface Structure of TiO2(011)-(2x1). *Physical Review Letters*, 93(3):036104, 2004.

[23] Xue-Qing Gong et al.˙ The 2x1 Reconstruction of the Rutile TiO2(011) Surface: A Combined Density Functional Theory, X-ray Diffraction, and Scanning Tunneling Microscopy Study. *Surface Science*, 603(1):138–144, 2009.

[24] X. Torrelles et al.˙ Geometric Structure of TiO2(011)(2x1). *Physical Review Letters*, 101(18):185501, 2008.

[25] Toshitaka Kubo, Hideo Orita, and Hisakazu Nozoye. Surface Structures of Rutile TiO2(011). *Journal of the American Chemical Society*, 129(34):1047410478, 2007.

[26] N S Beglitis et al. An STM analysis of the Stoichiometric TiO2(011) Surface. *Manuscript Under Preparation*, 2010.